



UNIwersytet  
Warszawski



---

Year: 2017

---

## Axiomatic Theories of Truth, Bounded Induction and Reflection Principles

Łełyk, Mateusz

Posted at The Institutional Repository of the University of Warsaw  
ReIn UW: <https://repozytorium.uw.edu.pl/handle/item/2266>  
Unique UUID of the publication: 2df4920d-6d2e-4450-a6e3-acff2cf278c6



The following work is licensed under a CC-BY - Attribution License.

Uniwersytet Warszawski  
Wydział Filozofii i Socjologii  
Instytut Filozofii

Mateusz Łełyk

# Axiomatic Theories of Truth, Bounded Induction and Reflection Principles

Rozprawa doktorska napisana pod kierunkiem  
dr. hab. Cezarego Cieślińskiego

Warszawa 2017

## CONTENTS

1. <i>Introduction</i> . . . . .	3
1.1 Structure of the Thesis . . . . .	5
2. <i>Formal Preliminaries</i> . . . . .	6
2.0.1 Peano Arithmetic . . . . .	6
2.0.2 Arithmetisation . . . . .	9
2.0.3 Models of PA . . . . .	23
2.0.4 Arithmetical Reflection . . . . .	27
3. <i>Axiomatic Theories of Truth</i> . . . . .	31
3.1 Definition and Motivations . . . . .	31
3.2 Strength of Axiomatic Theories of Truth . . . . .	38
3.2.1 Relative Truth Definability . . . . .	38
3.2.2 Model-theoretical Strength . . . . .	40
3.2.3 Proof-Theoretical Strength . . . . .	41
3.2.4 Relations between the three notions of strength . . . . .	42
3.3 Definitions of Axiomatic Theories of Truth in Study . . . . .	43
3.3.1 Classically Compositional Theories . . . . .	44
3.3.2 Non-Classically Compositional Theories of Truth . . . . .	45
3.3.3 Strength of Classically Compositional Theories . . . . .	50
3.3.4 Strength of Non-Classically Compositional Theories . . . . .	52
3.4 Reflection Principles . . . . .	54
3.5 Additional Axioms . . . . .	58
4. <i>Classically Compositional Truth Theories</i> . . . . .	62
4.1 Classical Compositional Truth with the $\Delta_0$ -induction . . . . .	62
4.2 Many Faces of $CT_0$ . . . . .	80
5. <i>Non-Classically Compositional Truth Theories</i> . . . . .	84
5.1 Bounded induction . . . . .	84
5.2 Disjunctive Correctness and Internal Induction . . . . .	96
5.2.1 Strong Kleene Case . . . . .	97
5.2.2 Weak Kleene Case . . . . .	109
5.3 Reflection principles . . . . .	116
6. <i>Summary: The Big Picture</i> . . . . .	121

## 1. INTRODUCTION

The main objective of this dissertation is to study properties of the notion of truth. Staying within the tradition initiated by Tarski, we take truth to be a property of sentences and apply formal tools in order to clarify and answer interesting us questions. To be more precise, our study consists of verifying how various properties of truth relate to each other. What is a *property of a notion*  $P$ ? For example non-emptiness, or dually, universality: the former applies to  $P$  if there is an object which satisfies  $P$ : the latter applies to  $P$  if every object is  $P$ . Not to invoke higher order objects (such as properties of properties), we may think of a property of  $P$  as of a linguistic law governing its use. Stating this a little bit more formally, they may be considered meaning postulates or *axioms* of our language that regulate the assertability conditions for sentences in which  $P$  occurs. In this dissertation we are interested in the case when  $P$  is the notion of truth and we investigate various *axiomatic theories of truth* which we take to model possible meaning postulates for this notion.

What are the properties of *truth*? Let us give some examples. Firstly, the notion of truth may be *self-applicable* (or *untyped*). In such a case we may make sense of sentences, such as

The sentence "The sentence "Snow is white" is true" is true.

When the notion of truth is defined only for sentences that themselves do not contain the truth predicate, we say that the notion of truth is *stratified* (or *typed*). In this dissertation we focus our attention exclusively on the axiomatic theories of *stratified* truth. Following Tarski, we agree that the minimal condition for a notion  $T$  to be treated as the notion of stratified truth for some language  $\mathcal{L}$  (an *object* language) is satisfying all sentences of the form:

$T("ϕ")$  if and only if  $ϕ$ ,

where " $ϕ$ " is a name of  $ϕ$  (i.e.  $ϕ$  is mentioned on the left-hand side of the above equivalence). What further properties of stratified truth we may consider? For example, *compositionality* is an obvious candidate: a notion of truth for a language  $\mathcal{L}$  is compositional (with respect to the set of basic connectives and quantifiers of  $\mathcal{L}$ ) if among axioms for  $T$  we have those saying how the truth of a compound sentence  $ϕ$  is related to the truth of its subformulae; e.g. the following sentence may be taken as axiom:

For every  $ϕ, ψ$ ,  $T("ϕ ∧ ψ")$  if and only if  $T("ϕ")$  and  $T("ψ")$ ,

where  $∧$  is a symbol for conjunction in  $\mathcal{L}$ . Similar conditions should then be stated for other connectives and quantifiers of  $\mathcal{L}$ . If negation  $¬$  is a connective of  $\mathcal{L}$ , we may also accept the following compositional axiom:

For every  $ϕ$ ,  $T("¬ϕ")$  if and only if not  $T("ϕ")$ . (NEG)

In our thesis we focus primarily on compositional theories, although not all theories we study admit NEG axiom: some of them will model a non-classical notion of truth.

Other properties the notion of truth might have are listed in the title of this dissertation. The first one is weak inductiveness, meaning that the notion of truth satisfies bounded induction. This has a rather technical flavour and comes from metamathematical investigations into Peano Arithmetic (as a first-order axiomatic theory; we introduce it properly in Chapter 2). In this theory, the fact that each set satisfies the induction principle is expressed via a scheme: for every formula of arithmetical language, we have a separate axiom stating that the set defined by this formula satisfies induction. However, in order to determine "the amount of arithmetic" needed to prove various combinatorial principles, one often studies subsystems of Peano Arithmetic, which admit an induction scheme restricted to a certain class of formulae. One of the most basic classes from which one usually starts consists solely of *bounded* formulae ( $\Delta_0$ ). The characteristic feature of these formulae is that in order to verify whether a given formula of this sort holds of an object  $a$  one needs to examine only a restricted fragment of the universe and which fragment should be examined can be read directly of the formula. In this dissertation we investigate which properties of truth follow when we assume that every bounded formula with the predicate  $T$  satisfies induction. The question is intriguing for at least two reasons. Firstly, with restricted possibilities of reasoning by induction, it is highly non-obvious whether one can invoke standard proof techniques (such as induction on the build-up of formulae, or induction on the length of proofs), which are normally used in order to demonstrate that the notion of truth has certain properties. Secondly, it is also highly unobvious what sort of sentences from the object language will be provable in such a restricted setting, with only a weakly inductive notion of truth at our disposal.

The current thesis contains original results on weakly inductive axiomatic truth theories, as described in the last paragraph. We prove that such basic axioms suffice to guarantee that the notion of truth is very well-behaved, meaning that it enjoys many further natural properties. Moreover, we show that this is the case for theories of both classically and non-classically compositional truth.

The last place on the list from the title of this dissertation is occupied by *reflection principles*. Intuitively, a reflection principle for a set of sentences  $X$  expresses the soundness of  $X$  with respect to some logic  $\mathcal{L}$ . In other words, the intuitive content of a reflection principle is that every sentence provable in a logic  $\mathcal{L}$  from a set  $X$  is true. As in the case of induction, this intuition is often expressed without the notion of truth. In such cases reflection for a language  $\mathcal{L}$  is typically presented as a schema encompassing all sentences of the form

$$\text{"if } \phi \text{ is provable from } X \text{ in logic } \mathcal{L}, \text{ then } \phi\text{"},$$

where  $\phi$  is a sentence of  $\mathcal{L}$ . What is important is that "being provable from a set  $X$  in logic  $\mathcal{L}$ " is expressed by a formula of  $\mathcal{L}$ . Having the truth predicate at our disposal, we may express directly the above intuition in a single sentence, for example,

$$\text{For every sentence } \phi \text{ of } \mathcal{L}, \text{ if } \phi \text{ is provable from } X \text{ in logic } \mathcal{L}, \text{ then } \phi \text{ is true.}$$

We shall distinguish *completeness* and *closure* reflection principles, depending on whether  $X$  contains the set of all true sentences (the latter type) or not (the former one). Let us note that

a *closure* reflection principle for a logic  $\mathcal{L}$  implies that the property of being true is preserved in reasoning in  $\mathcal{L}$ ; or, to put it differently, that the set of true sentences is *closed* under reasoning in  $\mathcal{L}$ . One of the main objectives of this thesis is to determine how such principles relate to the bounded induction for the truth predicate, if the latter is compositional. In particular, we shall ask whether bounded induction permits us to prove reflection principles, and if so, which forms of reflection become provable as soon as the bounded induction is added to our truth theory. Moreover, we shall ask whether this relation depends on how (classically or non-classically) compositional the truth predicate is. The thesis contains new results clarifying these dependencies.

Generally speaking, we are interested in the *strength* of the truth principles. We introduce three different formal explications of this notion, but are focused mainly on the *proof-theoretical one*: our aim is to characterize the sets of sentences of the object language  $\mathcal{L}$ , which can be deduced from various combinations of the truth properties we consider.

### 1.1 Structure of the Thesis

We shall focus on the case when the object language is the language of Peano Arithmetic (PA) and treat the latter as the object theory. That is why we introduce all the notions and facts from metamathematics of PA that we will need in this dissertation in the next chapter. Chapter 3 may be considered as a continuation of the above brief introduction: it begins with a presentation of motivations standing behind Axiomatic Theories of Truth, followed by the introduction of three formal explications of the notion of *strength* of an axiomatic theory of truth. Next, we define all the axiomatic truth (and satisfaction) theories that we will investigate and state all the known facts about their strength. This chapter ends with two subsections devoted to introducing various reflection principles and additional axioms with which we extend axiomatic theories. Chapter 4 is devoted to the proof of the Global Reflection principle in  $CT_0$  (i.e. the theory which contains basic compositional axioms for the truth predicate and induction for  $\Delta_0$  formulae of extended language). We show how to fix a gap in Kotlarski's original proof [28]. This is the first major original result of our dissertation. We end this chapter by showing the surprising equivalence between various extensions of  $CT^-$  (a basic, non-inductive compositional theory of truth) with the principles introduced in Chapter 3. The proof of this "Many Faces" Theorem is a combination of results due to Cieśliński ([7], [6], [4]), Enayat (unpublished) and the above mentioned result on provability of Global Reflection in  $CT_0$ . In Chapter 5 we show that this result does not transfer to the context of non-classically compositional theories of truth: in this case, analogous extensions of non-classically compositional theories generate not only different theories, but also theories which differ in arithmetical consequences. In particular, the intuitive difference emerges between *completeness* and *closure* reflection principles, the former being far *weaker* than the latter. All the results presented in Chapter 5 were, to our best knowledge, previously unknown. In Chapter 6 we summarize our findings.

## 2. FORMAL PRELIMINARIES

### 2.0.1 Peano Arithmetic

All theories we consider are formulated in an extension of *the first-order language of arithmetic*, in which  $\forall, \neg, \exists$  and  $=$  are basic logical constants (we treat  $\wedge, \rightarrow, \nabla, \neq$  as defined symbols). Let us start by introducing the non-logical constants of this language:

**Definition 1** (The language of arithmetic). *The language of arithmetic*, denoted  $\mathcal{L}_{\text{PA}}$ , is a first order language over signature  $\{0, 1, +, \cdot\}$ , where

1.  $0, 1$  are constants,
2.  $\cdot, +$  are binary function symbols.

The next definition introduces basic theory axiomatising the most important properties of symbols from our signature.

**Definition 2** (Robinson's Arithmetic  $Q$ ). The axioms of theory  $Q$  are the universal closures of the following formulae

1.  $x + 1 \neq 0$
2.  $(x + 1 = y + 1) \rightarrow x = y$
3.  $x \neq 0 \rightarrow \exists y (y + 1 = x)$
4.  $x + 0 = x$
5.  $x + (y + 1) = (x + y) + 1$
6.  $x \cdot 0 = 0$
7.  $x \cdot (y + 1) = x \cdot y + x$

**Remark 3** (Abbreviations). For the sake of convenience (more concretely, to reduce the number of shapes of atomic formulae), we decided not to include the relational symbol of ordering in  $\mathcal{L}_{\text{PA}}$  and axioms governing its use in  $Q$ . We shall treat  $x \leq y$  as an abbreviation of

$$\exists z (x + z = y),$$

where, for the sake of definiteness,  $z$  is the variable with least index different to  $x, y$ . As usual,  $x < y$  abbreviates  $x \leq y \wedge x \neq y$  and  $x \geq y, x > y$  abbreviate  $y \leq x$  and  $y < x$ , respectively. Moreover, we treat  $y = x - z$  as an abbreviation of

$$(x \geq z \wedge y + z = x) \vee (x < z \wedge y = 0).$$

Since PA proves that for each  $x, z$ , there exists  $y = x - z$  and is uniquely determined (see e.g. [19]), we will treat "  $-$  " as a binary function symbol.

**Remark 4.** The above is not what is usually defined as Robinson's Arithmetic, since the latter theory is formulated in the language with unary function symbol  $S$  (for successor), instead of constant 1 (sometimes also with a binary relational symbol  $\leq$  for ordering). The distinction is, however, purely notational, since by putting

$$S(x) := x + 1$$

our theory will prove all the usual axioms of  $Q$  (as defined e.g. in [19]<sup>1</sup>). *Vice versa*, traditional  $Q$  proves the axioms of our theory by interpreting 1 as  $S(0)$ . This is purely a matter of our taste that we prefer working with constant 1 instead of function  $S$ .

**Definition 5** (Peano Arithmetic). The axioms of *Peano Arithmetic* are all the axioms of  $Q$  together with universal closures of all formulae of the following form:

$$\phi(0) \wedge \forall x(\phi(x) \rightarrow \phi(x + 1)) \longrightarrow \forall x\phi(x),$$

where  $\phi$  is a formula of  $\mathcal{L}_{PA}$  and  $\phi(t)$  in the above formula denotes the result of substitution of a term  $t$  in the place of variable  $x$ . The universal closure of the above formula is also called *the instantiation of induction scheme with  $\phi$*  and denoted  $\text{Ind}(\phi)$ . If  $\mathcal{L}$  is any language, then  $\text{Ind}(\mathcal{L})$  denotes the set of axioms

$$\{\text{Ind}(\phi) \mid \phi \text{ is a formula of } \mathcal{L}\}.$$

If  $\text{Th}$  is an  $\mathcal{L}$  theory proving  $\text{Ind}(\mathcal{L})$ , then we call  $\text{Th}$  *fully inductive*.

**Proposition 6.** *PA proves that  $\leq$  is a discrete linear order, with 0 as the least element, 1 its immediate successor and such that for all  $x, y$*

$$\begin{aligned} x < y &\rightarrow \forall z(x + z < y + z) \\ x < y &\rightarrow \forall z(z \neq 0 \rightarrow x \cdot z < y \cdot z) \end{aligned}$$

Let us introduce a piece of notation:

**Definition 7.** Let  $\mathcal{L}$  be arbitrary first-order language.

1. If  $\phi(x_0, \dots, x_{n-1})$  is an arbitrary formula of  $\mathcal{L}$  and  $s_0, \dots, s_{n-1}$  are arbitrary terms of  $\mathcal{L}$ , then

$$\phi[s_0/x_0, \dots, s_{n-1}/x_{n-1}]^*$$

denotes the result of renaming bounded variables in  $\phi$  to make them disjoint from variables occurring in terms  $s_0, \dots, s_{n-1}$  and substituting  $s_i$  for  $x_i$ , for every  $i \leq n - 1$ .

2. If  $P$  is any  $n$ -ary predicate, then  $\mathcal{L} + P$  denotes the language extending  $\mathcal{L}$  with  $P$ . Instead of  $\mathcal{L}_{PA} + P$ , we shall be writing  $\mathcal{L}_P$ .
3. If  $\Theta$  is any formula of  $\mathcal{L} + P$ , and  $\phi(x_0, \dots, x_{n-1})$  is any formula having exactly  $x_0, \dots, x_{n-1}$  as free variables, then

$$\Theta[\phi(\bar{x})/P(\bar{x})]$$

is a formula defined recursively on the complexity of  $\Theta$ :

<sup>1</sup> Definition of  $Q$  in [19] includes also the definition of  $\leq$  in terms of addition. Obviously we can add it also in our case. For the definition of ordering see Remark 3.

(a) If  $\Theta = (s = t)$  for some terms  $s, t$ , then  $\Theta[\phi(\bar{x})/P(\bar{x})] = \Theta$ .

(b) If  $\Theta = P(s_0, \dots, s_{n-1})$  for some terms  $s_0, \dots, s_{n-1}$ , then

$$\Theta[\phi(\bar{x})/P(\bar{x})] = \phi[s_0/x_0, \dots, s_{n-1}/x_{n-1}]^*.$$

(c) If  $\Theta = \Phi \vee \Psi$  for some formulae  $\Phi, \Psi$  of  $\mathcal{L}_P$ , then

$$\Theta[\phi(\bar{x})/P(\bar{x})] = \Phi[\phi(\bar{x})/P(\bar{x})] \vee \Psi[\phi(\bar{x})/P(\bar{x})].$$

(d) If  $\Theta = \neg\Phi$  for some formula  $\Phi$  of  $\mathcal{L}_P$ , then  $\Theta[\phi(\bar{x})/P(\bar{x})] = \neg\Phi[\phi(\bar{x})/P(\bar{x})]$ .

(e) If  $\Theta = \exists x\Phi$  for some formula  $\Phi$  of  $\mathcal{L}_P$  and a variable  $x$ , then

$$\Theta[\phi(\bar{x})/P(\bar{x})] = \exists x\Phi[\phi(\bar{x})/P(\bar{x})].$$

Let  $\mathcal{L}$  be an arbitrary first-order language extending  $\mathcal{L}_{PA}$ . Each formula in the definition below is a formula of  $\mathcal{L}$  and Th is an arbitrary  $\mathcal{L}$  theory.

**Definition 8** (Arithmetical Hierarchy). We shall say that (the outermost occurrence of) quantifier  $\exists$  in the formula  $\exists x\phi$  is *bounded* if  $\phi$  is of the form  $(x < t \wedge \psi)$  for some formula  $\psi$  and some  $\mathcal{L}$  term  $t$ .

1. Formula  $\phi$  is in the class  $\Delta_0 (= \Sigma_0 = \Pi_0)$  if all quantifiers occurring in it are bounded.
2. Formula  $\phi$  is in the class  $\Sigma_{n+1}$  if for some  $k \in \omega$  it is of the form

$$\exists x_{i_1} \dots \exists x_{i_k} \psi$$

where  $\psi$  is in the class  $\Pi_n$  (we stipulate that if  $k = 0$ , then the above prefix is empty). We shall say that a formula  $\phi$  is  $\Sigma_n(\text{Th})$  if it is provably equivalent in Th to a formula in  $\Sigma_n$  class.

3. Formula  $\phi$  is in the class  $\Pi_{n+1}$  if for some  $k$  it is of the form

$$\forall x_{i_1} \dots \forall x_{i_k} \psi$$

where  $\psi$  is in the class  $\Sigma_n$  (we stipulate that if  $k = 0$ , then the above prefix is empty). We shall say that a formula  $\phi$  is of  $\Pi_n(\text{Th})$  class if it is provably in Th equivalent to a formula in  $\Pi_n$  class.

4. We shall say that a formula  $\phi$  is of  $\Delta_n(\text{Th})$  class if it is provably in Th equivalent to a formula of  $\Pi_n$  class and to a formula of  $\Sigma_n$  class.
5. If  $\phi(\bar{x})$  is any formula having exactly  $x_0, \dots, x_{n-1}$  as free variables, then  $\Delta_0(\phi(\bar{x}))$  denotes the class consisting of all formulae of the form

$$\Theta[\phi(\bar{x})/P(\bar{x})]$$

where  $P$  is an  $n$ -ary predicate not belonging to  $\mathcal{L}$  and  $\Theta$  is an arbitrary  $\Delta_0$  formula of  $\mathcal{L} + P$ .

**Definition 9** (Fragments of PA). For  $n \in \omega$ ,  $I\Sigma_n$  denotes the extension of  $Q$  with instantiations of induction scheme for formulae in  $\Sigma_n$  class. Instead of  $I\Sigma_0$ , we write  $I\Delta_0$ .

**Definition 10.** PAP is the theory formulated in  $\mathcal{L}_P$  whose only non-logical axioms are axioms of PA (i.e. we allow instantiations of induction scheme with *arithmetical* formulae only).  $I\Sigma_n(\mathcal{L}_P)$  is the theory extending PAP with instantiations of the induction scheme for  $\mathcal{L}_P$  formulae in  $\Sigma_n$  class.

### 2.0.2 Arithmetisation

We would like to develop a theory of syntax inside PA. To achieve this goal we will make a detour through set theory of hereditarily finite sets. In such a way the manner in which various notions are formalisable in PA will become clearer. This approach is implicit in [19]. Let us first introduce the relevant set theory - in our presentation we rely on [25] in which this theory was proved to be bi-interpretable with  $PA^2$ :

**Definition 11.** The theory  $ZF - \text{Inf}^*$  is a theory in the usual language of set theory containing the following axioms:

1. full schemata of Replacement and Separation;
2. usual Axioms of Sum, Extensionality, Empty Set, Pair, Powerset, Foundation
3. negation of the Axiom of Infinity
4. the following scheme of  $\in$ -induction:

$$\forall \bar{y} \left( \forall x \left( (\forall z \in x \phi(z)) \rightarrow \phi(x) \right) \rightarrow \forall x \phi(x) \right)$$

Within  $ZF - \text{Inf}^*$ , in the manner well-known from basic logic courses, one can talk about ordered pairs (of finite sets), functions (with finite domain), natural numbers (as Von-Neumann ordinals), (finite) sequences, terms, formulae, substitution functions and so on.

**Example 12.** Obviously, in  $ZF - \text{Inf}^*$  theory we cannot prove the existence of, for example, the set of all  $\mathcal{L}_{PA}$  terms but nevertheless we can define a formula  $\text{Term}_{\mathcal{L}_{PA}}(x)$  representing it. This is done in the standard manner for recursively defined objects: we put  $\text{termpos}_{\mathcal{L}_{PA}}(y, i)$  ("i-th position in a sequence  $y$  is occupied by a term") to be the following formula

$$\left( [y]_i = \underline{0} \vee [y]_i = \underline{1} \vee \exists j, k < i \left( ([y]_i = [y]_j \hat{\cdot} [y]_k) \vee ([y]_i = [y]_j \hat{\pm} [y]_k) \right) \right)$$

where  $\underline{0}, \underline{1}, \hat{+}, \hat{\cdot}$  are designated codes of 0, 1, + and  $\cdot$ , respectively, and  $[y]_l$  denote the projection on the  $l$ -th axis of  $y$ . Further define:

$$\begin{aligned} \text{termseq}'_{\mathcal{L}_{PA}}(y, x) &:= \text{Seq}(y) \wedge \forall i < \text{len}(y) \left( \text{termpos}_{\mathcal{L}_{PA}}(y, i) \right) \wedge \text{last}(y) = x, \\ \text{termseq}_{\mathcal{L}_{PA}}(y, x) &:= \text{termseq}'_{\mathcal{L}_{PA}}(y, x) \wedge \forall z < y \neg \text{termseq}'(z, x), \end{aligned}$$

<sup>2</sup> Although the hardest part, being a classical result due to Gödel, is not given there: to prove that PA interprets the relevant set theory one has to show that PA admits the exponential function (or coding of arbitrary sequences), which requires some trickery. For the proof of this fact, consult either [19], chapter I or [24], chapter 5

where  $\text{Seq}(x)$  means that  $x$  is a sequence, and  $\text{last}(y) = x$  means that  $x$  is the last element of the sequence  $y$ . Finally put

$$\text{Term}_{\mathcal{L}_{\text{PA}}}(x) := \exists y \text{termseq}_{\mathcal{L}_{\text{PA}}}(y, x).$$

It is a folklore result (carefully explained in [25]) that PA is really ZF – Inf\* "in disguise". More precisely, the theories are bi-interpretable (the latter notion being fully explained in [25]). In this dissertation, we will only use the fact that PA is at least as expressive as ZF – Inf\*. More concretely, we have the following

**Theorem 13.** *There exists a  $\Delta_0$  formula  $\phi_{\in}(x, y) \in \mathcal{L}_{\text{PA}}$ , such that for every axiom  $\Phi$  of ZF – Inf\* we have*

$$\text{PA} \vdash \Phi[\phi_{\in}(x, y)/x \in y]$$

where  $\Phi[\phi_{\in}(x, y)/x \in y]$  is as explained in 7.

In other words (not to be explained here in full generality), there is a direct and unrelativised interpretation of  $\text{ZF}^- + \text{Inf}^*$  in PA. In the context of arithmetic, we shall abbreviate the formula  $\phi_{\in}(x, y)$  with  $x \in y$ . Let us note a consequence of the above theorem:

**Proposition 14** (Extensionality in PA). *The following sentence is provable in PA:*

$$\forall x, y \left( x = z \equiv (\forall z (z \in x \equiv z \in y)) \right)$$

We will use the above proposition when introducing new definitions in PA: it states that, as in set-theory, to define an object it is enough to say which elements it contains. In the definition below, we state usual set-theoretical definitions *inside* PA using the above-mentioned interpretation.

**Definition 15** (PA).

1. The ordered pair of  $x, y$  is  $\{\{x, y\}\{x\}\}$ . The ordered pair of  $x, y$  is denoted with  $\langle x, y \rangle$ .
2.  $x$  is a function if its sole elements are ordered pairs and for all  $y, z, z'$ , if  $\langle y, z \rangle \in x$  and  $\langle y, z' \rangle \in x$ , then  $y' = y$ .
3. If  $x$  is any function, then  $\text{dom}(x)$  and  $\text{im}(x)$  denote the *domain* of function  $x$  and the *image* of function  $x$ , respectively; i.e.

$$y \in \text{dom}(x) \equiv \exists z < x \langle y, z \rangle \in x$$

and

$$y \in \text{im}(x) \equiv \exists z < x \langle z, y \rangle \in x$$

4. A sequence is any function whose domain is downward closed; i.e.  $y$  is a sequence if  $y$  is a function and for all  $x$ , if  $x \in \text{dom}(y)$ , then for all  $z < x$ ,  $z \in \text{dom}(y)$ . The fact that  $y$  is a sequence is denoted with  $\text{Seq}(y)$ . If  $x_0, \dots, x_{n-1}$  are any elements, then

$$\langle x_0, \dots, x_{n-1} \rangle$$

denotes the (unique)  $n$ -elementary sequence with  $x_0, x_1, \dots$  as first, second, ... elements. The *length* of the sequence  $x$  is the number  $\max(\text{dom}(x)) + 1$  (where  $\max(y)$  is the maximal element of a set  $y$ ). Each number smaller than the length of  $x$  is called a *position* in  $x$ .  $\text{last}(x)$  denotes the *last* element of sequence  $x$ ; i.e. the element occupying position  $\text{len}(x) - 1$ .

5. If  $c$  is any sequence, then  $\bigcup c$  denotes the generalised sum of  $c$ ; i.e.

$$x \in \bigcup c \equiv \exists z \in \text{im}(c) \ x \in z.$$

If  $c_0, \dots, c_a$  is an indexed family of sets then we also use  $\bigcup_{i \leq a} c_i$  with the usual reading. If  $a, b$  are two sets then  $a \cup b$  denotes  $\bigcup \{a, b\}$ .

The following fact follows by the inspection of the usual definitions and the fact that  $\phi_{\in}(x, y)$  is  $\Delta_0(\text{PA})$ .

**Fact 16.**  $y = \langle x, y \rangle, y \in \text{dom}(x), y \in \text{im}(x), \text{Seq}(x), x \in \bigcup y$  are formulae of the class  $\Delta_0(\text{PA})$ .

All the below definitions are stated within PA.

**Definition 17** (Syntax in PA).

1.  $x$  is a *variable* (denoted  $\text{Var}(x)$ ) if for some  $y < x, x = \langle 0, y \rangle$ .
2.  $\text{termseq}_{\mathcal{L}_{\text{PA}}}(y, x)$  ( $y$  is the generating sequence of  $x$ ) and  $\text{Term}_{\mathcal{L}_{\text{PA}}}(x)$  are as defined in Example 12. Since all languages we consider contains the same terms we skip the reference to  $\mathcal{L}_{\text{PA}}$  in the subscript and write simply  $\text{termseq}(y, x)$  and  $\text{Term}(x)$ . Term  $x$  is *closed* (denoted  $\text{CTerm}_{\mathcal{L}_{\text{PA}}}(x)$ ) if no variable occurs in the generating sequence of  $x$ .
3.  $y$  is the *numeral naming*  $x$ , if  $y$  is equal to the term of the form

$$\underbrace{(1 + (1 + (\dots + (1 + 0) \dots))}_{x \text{ times } 1}$$

if  $x > 0$  or  $\underline{0}$  if  $x = 0$ . The numeral naming  $x$  will be denoted  $\underline{x}$ .

4.  $\ulcorner = \urcorner, \ulcorner \vee \urcorner, \ulcorner \exists \urcorner, \ulcorner \neg \urcorner, \ulcorner (\urcorner, \urcorner) \urcorner$  are any distinct numbers that are not terms<sup>3</sup>
5.  $y$  is the *generating sequence* of  $x$  if
  - (a)  $x$  is the last element of  $y$
  - (b) for every position  $i$  in  $y, [y]_i$  either
    - i. is equal to  $\langle \ulcorner (\urcorner, s, \ulcorner = \urcorner, t, \urcorner) \urcorner \rangle$  for some terms  $s, t$  or
    - ii. is equal to  $\langle \ulcorner (\urcorner, \phi, \ulcorner \vee \urcorner, \psi, \urcorner) \urcorner \rangle$  for some  $\phi, \psi$  occurring in  $y$  on positions strictly smaller than  $i$  or
    - iii. is equal to  $\langle \ulcorner (\urcorner, \neg, \psi, \urcorner) \urcorner \rangle$  for some  $\psi$  occurring in  $y$  on position strictly smaller than  $i$  or

<sup>3</sup> One may define them as (respectively) the first, the second, ... non-interesting numbers.

iv. is equal to  $\langle \ulcorner (\ulcorner, \ulcorner \exists \urcorner, v, \psi, \ulcorner) \urcorner \rangle$  where  $v$  is a variable and  $\psi$  occurs in  $y$  on position strictly smaller than  $i$

(c)  $y$  is the least number satisfying both the above conditions.

If  $y$  is the generating sequence of  $x$ , we denote it with  $\text{formseq}_{\mathcal{L}_{\text{PA}}}(x)$ .  $x$  is a *formula* of  $\mathcal{L}_{\text{PA}}$  (denoted  $\text{Form}_{\mathcal{L}_{\text{PA}}}(x)$ ) if there exists the generating sequence of  $x$ .

6.  $y$  is a *subformula* of  $x$  if it both  $x, y$  are formulae and  $y$  occurs in the generating sequence of  $x$ . If  $y$  is a subformula of  $x$ , then we denote it with  $\text{Subf}(y, x)$  or  $y \trianglelefteq x$ .
7. An *assignment* is any function whose domain consists solely of variables. If  $y$  is an assignment we denote it with  $y \in \text{Asn}$ .
8.  $\Sigma_x(y)$  denotes that  $y$  is a formula in  $\Sigma_x$  form (defined by inspection on the generating sequence of  $x$ ).  $\Pi_x(y)$  has analogous meaning.

**Fact 18.**  $\text{Var}(x)$ ,  $\text{termseq}(y, x)$ ,  $\text{formseq}_{\mathcal{L}_{\text{PA}}}(y, x)$ ,  $y \in \text{Asn}$  are  $\Delta_0(\text{PA})$  formulae.  $\text{Term}(x)$ ,  $\text{CTerm}_{\mathcal{L}_{\text{PA}}}(x)$ ,  $y = \underline{x}$ ,  $\text{Form}_{\mathcal{L}_{\text{PA}}}(x)$ ,  $\text{Subf}(y, x)$  and  $\Sigma_x(y)$  are  $\Delta_1(\text{PA})$  formulae.

**Convention 1.** Instead of  $\langle 0, k \rangle$  we shall be writing  $v_k$ , to denote the  $k$ -th variable. We will use also  $w, w_0, w_1, \dots$  to denote variables.

The next definitions show the concrete way of formalising the notion of the *syntactic tree* of a formula and an *occurrence* of a term or a subformula inside another term, or formula. These formalisations are developed more carefully, since they will play an important role in some of our arguments, and there seems to be no well-established way of operating them within PA (or ZF).

**Definition 19** (Tree). If  $\sigma$  and  $\tau$  are two sequences of 0, 1, then we say that  $\sigma$  is a *prefix* of  $\tau$ , if  $\text{len}(\sigma) \leq \text{len}(\tau)$  and for every position  $i$  in  $\sigma$ ,  $\sigma(i) = \tau(i)$ . If  $A$  is a set of finite 0, 1 sequences then  $\text{Pref}(A)$  denotes the set of all prefixes of elements of  $A$ . We shall denote the one element sequence consisting of 0 or 1 by 0 or 1, respectively. A *binary tree* is a set of sequences of 0, 1 closed under prefixes (including the empty prefix  $\varepsilon$ ). If  $\sigma$  is a sequence and  $A$  is a tree, then

$$\sigma \frown A := \text{Pref}(\{\sigma \frown \tau \mid \tau \in A\})$$

If  $A$  is a tree and  $\sigma, \tau \in A$ , then we define  $\sigma \prec_A \tau$  if and only

1. either  $\tau$  is a proper prefix of  $\sigma$  or
2.  $\tau$  is not a proper prefix of  $\sigma$  and if  $n$  is the least number such that  $\sigma(n) \neq \tau(n)$ , then  $\sigma(n) < \tau(n)$ .

So defined ordering  $\prec_A$  is linear.

**Definition 20** (Syntactic tree of a formula; PA). If  $s$  is a term, then the *syntactic tree* of  $s$  is a pair  $\langle A^s, l^s \rangle$  where

1.  $A^s$  is a binary tree

2.  $l^s$  is a function  $A^s \rightarrow \text{Terms}(s)$

defined by induction on the structure of  $s$ :

1. if  $s \in \text{Var} \cup \{0, 1\}$  then  $A^s = \{\varepsilon\}$ ,  $l^s(\varepsilon) = s$
2. if  $s = t \odot u$ , where  $\odot \in \{\cdot, +\}$ , then  $A^s = \{\varepsilon\} \cup 0 \frown A^t \cup 1 \frown A^u$ ,  $l^s(\varepsilon) = s$  and
  - (a) for every  $\sigma \in A^t$ ,  $l^s(0 \frown \sigma) = l^t(\sigma)$ ,
  - (b) for every  $\sigma \in A^u$ ,  $l^s(1 \frown \sigma) = l^u(\sigma)$ .

If  $\phi$  is a formula, then the *syntactic tree* of  $\phi$  is a pair  $\langle A^\phi, l^\phi \rangle$  where

1.  $A^\phi$  is a binary tree, called the *full skeleton* of  $\phi$ .
2.  $l^\phi$  is a function  $A^\phi \rightarrow \text{Subf}(\phi) \cup \text{Terms}(\phi)$ .

defined by induction on the structure of  $\phi$ :

1. if  $\phi = (s = t)$  for some terms  $s, t$  then  $A^\phi = \{\varepsilon\} \cup 0 \frown A^s \cup 1 \frown A^t$ ,  $l^\phi(\varepsilon) = \phi$  and
  - (a) for every  $\sigma \in A^s$ ,  $l^\phi(0 \frown \sigma) = l^s(\sigma)$ ,
  - (b) for every  $\sigma \in A^t$ ,  $l^\phi(1 \frown \sigma) = l^t(\sigma)$ .
2. if  $\phi = \neg\psi$  or  $\phi = \exists v\psi$ , then  $A^\phi = \{\varepsilon\} \cup 0 \frown A^\psi$ ,  $l^\phi(\varepsilon) = \psi$  and for every  $\sigma \in A^\psi$ ,  $l^\phi(0 \frown \sigma) = l^\psi(\sigma)$ .
3. if  $\phi = \psi_0 \vee \psi_1$  for some formulae  $\psi_0, \psi_1$ , then  $A^\phi = \{\varepsilon\} \cup 0 \frown A^{\psi_0} \cup 1 \frown A^{\psi_1}$ ,  $l^\phi(\varepsilon) = \phi$  and
  - (a) for every  $\sigma \in A^{\psi_0}$ ,  $l^\phi(0 \frown \sigma) = l^{\psi_0}(\sigma)$ ,
  - (b) for every  $\sigma \in A^{\psi_1}$ ,  $l^\phi(1 \frown \sigma) = l^{\psi_1}(\sigma)$ .

The above definition gives a perspicuous presentation of both propositional and term structure of a formula. We might need also one more notion making precise the idea of a purely *propositional* structure of a formula. If  $\phi$  is a formula, then the *reduced syntactic tree* of  $\phi$  is a pair  $\langle A^\phi, l^\phi \rangle$  satisfying the above definition of a syntactic tree in which we change condition 1 to 1':

1' if  $\phi = (s = t)$  for some terms  $s, t$  then  $A^\phi = \{\varepsilon\}$  and  $l^\phi(\varepsilon) = \phi$ .

If  $\langle A^\phi, l^\phi \rangle$  is the reduced syntactic tree of  $\phi$ , then  $A^\phi$  is called the *reduced skeleton*, denoted  $\text{Skel}(\phi)$ .

**Definition 21** (Occurrences; PA). Let  $\phi$  be formula.

1. If  $\sigma \in A^\phi$ , then  $l_\sigma^\phi$  denotes the restriction of  $l^\phi$  to the set of prefixes of  $\sigma$ .
2. If  $s$  is an arbitrary term, then we say that  $s$  *occurs* in  $\phi$  if there exists  $\sigma \in A^\phi$  such that  $l^\phi(\sigma) = s$ . If  $s$  occurs in  $\phi$  then if  $\sigma$  is such that  $l^\phi(\sigma) = s$ , then  $l_\sigma^\phi$  is called an *occurrence* of  $s$  in  $\phi$ . The notion of occurrence of a subformula of  $\phi$  is defined analogously.

3. Let  $l_\sigma^\phi$  and  $l_\tau^\phi$  be two occurrences of terms in  $\phi$ . We define  $l_\sigma^\phi \prec_\phi l_\tau^\phi$  if  $\sigma \prec_{A^\phi} \tau$ .
4.  $\text{Var}(\phi)$  denotes the set of variables which occur in  $\phi$  ( $\text{Var}(t)$  has an analogous meaning for a term  $t$ ). If  $l_\sigma^\phi$  is an occurrence of a variable in  $\phi$  then we say that  $l_\sigma^\phi$  is a *bounded* occurrence of variable  $v$  if for some prefix  $\tau$  of  $\sigma$  and some  $\psi \in \text{Subf}(\phi)$  we have  $l_\tau^\phi = \exists v\psi$ . We say that  $l_\sigma^\phi$  is a *free* occurrence of variable  $v$  if and only if  $l_\sigma^\phi$  is not bounded. The sets of variables which have free (bounded) occurrence in  $\phi$  will be denoted by  $\text{FV}(\phi)$  ( $\text{BV}(\phi)$ ).

**Remark 22.** Let us stress that provably in PA, for every  $\phi$   $\text{Var}(\phi)$   $\text{FV}(\phi)$ ,  $\text{BV}(\phi)$  exists as *finite* sets.

**Definition 23.**

1.  $\text{Form}_{\mathcal{L}_{\text{PA}}}^{\leq 1}(\phi)$  abbreviates  $(\text{Form}_{\mathcal{L}_{\text{PA}}}(\phi) \wedge \exists v \text{FV}(\phi) \subseteq \{v\})$ .
2.  $\text{Sent}_{\mathcal{L}_{\text{PA}}}(\phi)$  abbreviates  $(\text{Form}_{\mathcal{L}_{\text{PA}}}(\phi) \wedge \text{FV}(\phi) = \emptyset)$ .

**Definition 24** (PA). The complexity of  $\phi$  is the maximal length of a path in the *reduced* syntactic tree of  $x$ . We take  $\text{Compl}(\phi) = y$  to abbreviate that the complexity of  $\phi$  is equal to  $y$ .

**Definition 25** (Assignments and Substitution; PA).

1. Let  $\phi$  be a formula and  $\alpha$  an assignment. By  $\phi[\alpha]$ , we mean the formula resulting from  $\phi$  by substituting numerals for free variables of  $\phi$  such that  $\underline{x}$  is substituted for  $v_k$  if and only if  $\alpha(v_k) = x$ .
2. If some variables  $v_{i_0}, \dots, v_{i_k}$  are specified and  $t_{i_0}, \dots, t_{i_k}$  are any terms then we use

$$\phi(t_{i_0}/v_{i_0}, \dots, t_{i_k}/v_{i_k})$$

to denote the result of formal substitution of  $t_{i_j}$  for  $v_{i_j}$  in  $\phi$  for every  $j \leq k$ . Sometimes, we also abbreviate this with  $\phi(t_{i_0}, \dots, t_{i_k})$ , implicitly assuming that the numeral denoting an object with index  $i_k$  is substituted for variable with the same index. If  $\phi$  has at most one free variable, then we usually denote the result of the substitution of  $t$  for the unique free variable of  $\phi$  by simply  $\phi(t)$ .

**Example 26** (PA). Let  $\phi = \ulcorner \forall v_1 (v_0 + v_1 = v_3) \urcorner$  and let  $\alpha$  map  $v_0$  to 3 and  $v_3$  to 0. Then

$$\phi[\alpha] = \ulcorner \forall v_1 ((1 + (1 + (1 + 0))) + v_1 = 0) \urcorner$$

**Remark 27** (PA). Let us observe that if  $\phi, \psi$  are two formulae such that for some assignments  $\alpha, \beta$ ,

$$\phi[\alpha] = \psi[\beta]$$

then the reduced skeletons of  $\phi$  and  $\psi$  are the same.

**Definition 28.** If  $\phi$  is a formula, then  $\ulcorner \phi \urcorner$  denotes either the unique code of  $\phi$ , obtained via the translation of  $\text{ZF} - \text{Inf}^+$  in PA, or the numeral denoting this code (depending on whether  $\ulcorner \phi \urcorner$  is an element of the model or occurs in a formula).

The above definition describes the unique context of using  $\ulcorner \cdot \urcorner$  function that we will maintain in this dissertation. In all the other situations when talking about syntax we adopt notational conventions familiar from defining syntactical notions in ZF. Convention below highlights the most important points:

**Convention 2.** If  $\phi(x_0, \dots, x_n, y)$  is, provably in PA, a functional formula then we often treat it as a new function symbol  $\phi(x_0, \dots, x_n)$ . In particular, if  $\Psi(x)$  is any formula then for any  $x_0, \dots, x_n$  we write  $\Psi(\phi(x_0, \dots, x_n))$  instead of  $\exists y(\phi(x_0, \dots, x_n, y) \wedge \Psi(y))$ . Adopting this convention, we use the following abbreviations:

1.  $s + t$  denotes the term (of arithmetised language) formed by concatenating term  $s$ , symbol  $+$  and term  $t$ .
2.  $s \cdot t$  denotes the term (of arithmetised language) formed by concatenating term  $s$ , symbol  $\cdot$  and term  $t$ .
3.  $s = t$  denotes the atomic formula (of arithmetised language) formed by concatenating term  $s$ , symbol  $=$  and term  $t$ .
4.  $\phi \vee \psi$  denotes the formula (of arithmetised language) formed by concatenating formula  $\phi$ , symbol  $\vee$  and formula  $\psi$ .
5.  $\neg\phi$  denotes the formula (of arithmetised language) formed by concatenating symbol  $\neg$  with the formula  $\phi$ .
6.  $\exists v\phi$  denotes the formula (of arithmetised language) formed by concatenating symbol  $\exists$ , variable  $v$  and the formula  $\phi$ .

To avoid confusion, variables  $s$  and  $t$  will be used only in the context of terms, so, for example,  $s + t$  is never to be treated as the sum of two numbers  $s$  and  $t$ , but as a *term* of the arithmetised language. We use  $\phi, \psi, v$  in similar fashion. Moreover, we will use metavariables in order to simplify formulae:

1. by writing  $\forall v, \exists w$ , we implicitly quantify over variables;
2. by writing  $\forall s(\bar{v}), \exists t(\bar{w})$ , we implicitly quantify over terms;
3. by writing  $\forall s, \exists t$ , we implicitly quantify over closed terms;
4. by writing  $\forall \phi(\bar{v}), \exists \psi(\bar{w}), \dots$ , we implicitly quantify over arithmetical formulae;
5. by writing  $\forall \phi(v), \exists \psi(v), \dots$ , we implicitly quantify over arithmetical formulae with at most one free variable  $v$ ;
6. by writing  $\forall \phi, \exists \psi, \dots$ , we implicitly quantify over arithmetical sentences;
7. by writing  $\forall \alpha, \exists \beta, \dots$ , we implicitly quantify over assignments;
8. by writing  $\forall \sigma, \exists \tau, \dots$ , we implicitly quantify over sequences (sometimes we specify further what kind of sequences we want to quantify over).

Sometimes, especially in the context of  $<$ , we will explicitly make use of the fact that we treat formulae, terms, sequences as particular kind of numbers and write things such as

$$x < \phi \wedge \psi$$

to mean that *the number*  $x$  is smaller than *the number*  $\phi \wedge \psi$ . However, as already indicated, we never use  $s + t$  ( $s \cdot t$ ) to denote the sum (product) of two numbers.

**Convention 3.** We shall stretch the above convention even further by making  $\forall\phi(v)$ ,  $\forall\phi(\bar{x})$ ,  $\forall\phi$  etc. *case-sensitive*: in the course of our thesis we shall define within PA various extensions of  $\mathcal{L}_{\text{PA}}$  and while working with each of them, we will assume that  $\phi$ ,  $\phi(\bar{x})$ , etc. range over syntactical objects of the appropriate extension. Sentences in which  $\phi$  etc. range over syntactical objects of  $\mathcal{L}$  will be called  $\mathcal{L}$ -variants. For example, the  $\mathcal{L}$ -variant of sentence

$$\forall\phi T(\phi)$$

is

$$\forall x (\text{Sent}_{\mathcal{L}}(x) \rightarrow T(x)).$$

By default (under the absence of additional specifications),  $\phi$ ,  $\phi(\bar{x})$  etc. range over syntactical objects of  $\mathcal{L}_{\text{PA}}$ .

**Definition 29 (PA).**  $\text{Ind}(v, \phi)$  denotes the instantiation of an induction axiom by  $\phi$  w.r.t.  $v$ , i.e. the sentence

$$\forall\bar{w} \left( (\phi[0/v] \wedge \forall v (\phi \rightarrow \phi[v + 1/v])) \rightarrow \forall v \phi \right)$$

where  $\bar{w}$  contains all the free variables of  $\phi$ , except of  $v$ .

Note that the above definition concerns sentences of arithmetised language.

**Remark 30.** Our convention, while in general helping to keep our formulae readable, sometimes might hide delicate points. For example, if  $\Psi$  is a  $\mathcal{L}_T$  formula of class  $\Delta_0$ , then

$$\Phi := \forall\phi, \psi < z (\Psi(\phi \wedge \psi))$$

looks like a formula of class  $\Delta_0$ , but is really  $\Delta_1(\text{PAT})$ , since, using our convention, it is equal to the formula

$$\forall x, y < z (\text{Form}_{\mathcal{L}_{\text{PA}}}(x) \wedge \text{Form}_{\mathcal{L}_{\text{PA}}}(y) \rightarrow \Psi(x \wedge y))$$

and

1.  $\text{Form}_{\mathcal{L}_{\text{PA}}}(x)$  is a  $\Delta_1(\text{PA})$  formula and
2.  $\Psi(x \wedge y)$  is really a  $\Delta_1(\text{PAT})$  formula.

This is particularly important, since one of the main focus of our thesis are truth theories with  $\Delta_0$  induction only. Hence, at some points it will demand justification that we are allowed to use induction axiom for a formula. We will indicate the way of solving this problem in the next subsection.

**Convention 4.** If  $\Psi$  is a formula,  $x$ - a variable and  $\phi$  a code of a formula with at most one free variable, then  $\Psi(\phi(\underline{x}))$  denotes the formula with free variable  $x$  of the form

$$\exists y, z \ (y = \underline{x} \wedge z = \phi(y) \wedge \Psi(z))$$

**Example 31.** Let us show a typical example of applying the above convention: an  $\mathcal{L}_T$  formula

$$\forall v \forall \phi(v) \exists x T(\phi(\underline{x}))$$

is a short for

$$\forall v \forall \phi(v) \exists x \left( \exists y, z (y = \underline{x} \wedge z = \phi(y) \wedge T(z)) \right)$$

In what follows, we take Th to be any fully inductive theory in  $\mathcal{L}_P$  with recursively enumerable set of axioms and extending PA.

**Definition 32** (The value of a term; Th). If  $t$  is any term and  $\alpha$  an assignment such that  $\text{dom}(\alpha) \subseteq \text{Var}(t)$ , then  $(t)_\alpha^\circ$  denotes the *value of  $t$  under  $\alpha$* , defined as in [24]. If  $t$  is a closed term, then  $t^\circ$  abbreviates  $(t)_\varepsilon^\circ$ , where  $\varepsilon$  is the empty valuation.

**Definition 33** (Partial Truth Predicates; Th).  $\text{Sat}_{\Sigma_n}(x, y)$  denotes the standard satisfaction predicate for  $\Sigma_n$  formulae of  $\mathcal{L}_P$ , defined (for Th = PA) as in [19] or [24]. We note that  $\text{Sat}_{\Sigma_n}(x, y)$  is a  $\Sigma_n(\text{Th})$  formula. Using this predicate, we define the truth predicate for  $\Sigma_n$  sentences of  $\mathcal{L}_P$  by putting

$$\text{Tr}_{\Sigma_n}(\phi) := \text{Sat}_{\Sigma_n}(\phi, \varepsilon)$$

where  $\varepsilon$  denotes the empty valuation.  $\text{Tr}_{\Sigma_n}(x)$  is again a  $\Sigma_n$  formula of  $\mathcal{L}_P$ .

**Proposition 34** ("It's snowing"-It's snowing lemma, [19] Corollary 1.76). *For every sentence  $\phi$  of  $\Sigma_k$  class  $I\Sigma_1$  proves*

$$\text{Tr}_{\Sigma_k}(\ulcorner \phi \urcorner) \equiv \phi.$$

Moreover the above can be arithmetised, so for every  $k$

$$\text{PA} \vdash \forall \phi \text{Pr}_{I\Sigma_1}(\text{Tr}_{\Sigma_k}(\underline{\phi}) \equiv \phi)$$

The above-defined partial truth predicates work only for formulae in  $\Sigma_n$  forms. At some points we will need ones defined for arbitrary sentences, which are delivered by the following definition:

**Definition 35** (Partial Truth Predicates for Arbitrary Sentences). Let  $\text{Tr}_n(x)$  denote the partial truth predicate for sentences of complexity  $\leq n$ , defined as in [38]

**Proposition 36** (Pudlak, [38]). *For every formula  $\phi(x_0, \dots, x_k)$  of complexity  $\leq n$ , PA proves*

$$\forall t_0, \dots, t_k \ (\text{Tr}_n(\ulcorner \phi(t_0, \dots, t_k) \urcorner) \equiv \phi(t_0^\circ, \dots, t_k^\circ))$$

Definition below (for Th = PA) is taken from [19].

**Definition 37** (Definable sets; Th). We say that  $a$  is a  $\Sigma_n$  set if  $a$  is a  $\Sigma_n$  formula of  $\mathcal{L}_P$  with exactly one free variable  $v_0$ .  $\Pi_n$  sets are defined analogously. If  $a$  is a  $\Sigma_n$  set, then we define

$$x \in_{\Sigma_n} b := \text{Sat}_{\Sigma_n}(b, [x])$$

where  $[x]$  denotes the function  $v_0 \mapsto x$ .  $\in_{\Pi_n}$  is defined analogously. A  $\Delta_n$  set is a pair  $\langle a, b \rangle$  such that  $a$  is a  $\Sigma_n$  formula,  $b$  is a  $\Pi_n$  formula and we have

$$\forall x (x \in_{\Sigma_n} a \equiv x \in_{\Pi_n} b)$$

If  $x = \langle a, b \rangle$  is a  $\Delta_n$  set, we define:

1.  $z \in_n x := z \in_{\Sigma_n} a$
2.  $x^2$  is the  $\Sigma_n$  formula equivalent to

$$\exists v_1, v_2 (v_0 = \langle v_1, v_2 \rangle \wedge v_1 \in_n x \wedge v_2 \in_n x)$$

Moreover if  $y = \langle c, d \rangle$  is another  $\Delta_n$  set, then we define

$$x \subseteq y := \forall z (z \in_{\Sigma_n} a \rightarrow z \in_{\Pi_n} d)$$

**Definition 38** (Th).  $\text{Th}'$  is a  $\Delta_n$  arithmetical theory if  $\text{Th}'$  is a  $\Delta_n$  set of sentences; i.e.  $\text{Th}'$  is such that

$$\forall x (x \in_n \text{Th}' \rightarrow \text{Sent}_{\mathcal{L}_{\text{PA}}}(x))$$

If  $\text{Th}'$  is a formula such that for some  $n$   $\text{Th} \vdash \ulcorner \text{Th}' \urcorner$  is a  $\Delta_n$  arithmetical theory", then we shall call  $\text{Th}'$  *Th provable theory*.

**Definition 39** (Provability in Classical Propositional Calculus; Th). Let  $\text{Th}'$  be a  $\Delta_n$  arithmetical theory. Formula  $\psi$  is *provable in Classical Propositional Calculus from  $\text{Th}'$*  if there exists a sequence  $\sigma$  such that  $\text{last}(\sigma) = \psi$  and for every position  $k$  in  $\sigma$  one of the following holds:

1.  $\sigma(k) \in_n \text{Th}'$  or it is a formula of one of the following kinds
  - (a)  $\phi \rightarrow (\theta \rightarrow \phi)$
  - (b)  $(\phi \rightarrow (\theta \rightarrow \chi)) \rightarrow ((\phi \rightarrow \theta) \rightarrow (\phi \rightarrow \chi))$
  - (c)  $\phi \rightarrow \phi \vee \theta$
  - (d)  $\theta \rightarrow \phi \vee \theta$
  - (e)  $(\phi \rightarrow \chi) \rightarrow ((\theta \rightarrow \chi) \rightarrow (\phi \vee \theta \rightarrow \chi))$
  - (f)  $\phi \wedge \theta \rightarrow \theta$
  - (g)  $\phi \wedge \theta \rightarrow \phi$
  - (h)  $\phi \rightarrow (\theta \rightarrow \phi \wedge \theta)$
  - (i)  $\neg\neg\phi \rightarrow \phi$
2. for some  $i, j < k$ ,  $\sigma(j) = \sigma(i) \rightarrow \sigma(k)$ .

$\text{Pr}_{\text{CPC}}^{\text{Th}'}(\psi)$  abbreviates the formula constructed in the above way.  $\text{Pr}_{\text{CPC}}(\phi)$  abbreviates  $\text{Pr}_{\text{CPC}}^{\emptyset}(\phi)$ .

We proceed to arithmetising provability in First Order Logic. Two technical notions will be needed:

**Definition 40 (Th).** Arithmetical formula  $\psi$  is a *generalisation* of a formula  $\phi$  if for some  $x$  and some variables  $v_0, \dots, v_x$ ,

$$\psi = \forall v_0 \dots \forall v_x \phi$$

We say that a term  $t$  can be substituted without clashes in a formula  $\phi$  for a variable  $x$  if for every free occurrence  $l_\sigma^\phi$  of  $x$  in  $\phi$ , for no  $v \in \text{Var}(t)$ , no  $\tau \in \text{Suff}(\sigma)$  and no formula  $\psi$  it holds that  $l_\tau^\phi = \exists v \psi$ .

The above definition of the relation " $t$  can be substituted without clashes for  $x$  in  $\phi$ " is as usual for First-Order Logic: we make sure that no variable of  $t$  will become bounded after the substitution.

**Definition 41 (Provability in  $\text{Th}'$ ; Th).** If  $\text{Th}'$  is a  $\Delta_n$  arithmetical theory and  $\psi$  is any arithmetical formula, then  $\psi$  is *provable* in  $\text{Th}'$  if there exists a sequence of formulae  $\sigma$  such that  $\text{last}(\sigma) = \psi$  and for every position  $k$  in  $\sigma$  one of the following holds:

1.  $\sigma(k)$  is an axiom of First-Order Logic, i.e. a generalisation of a formula of the following kind<sup>4</sup>:
  - (a) an axiom of Classical Propositional Calculus;
  - (b)  $\forall x \phi \rightarrow \phi[t/x]$ , where  $t$  can be substituted for  $x$ ;
  - (c)  $\forall x(\phi \rightarrow \psi) \rightarrow (\forall x \phi \rightarrow \forall x \psi)$ , for all  $\phi, \psi$ ;
  - (d)  $\phi \rightarrow \forall x \phi$  where  $x \notin \text{FV}(\phi)$ , for all  $\phi$ ;
  - (e)  $x = x$  for all  $x \in \text{Var}$ ;
  - (f)  $x = y \rightarrow \phi \equiv \phi'$  for all  $\phi$  where  $\phi'$  is obtained by replacing some or zero occurrences of  $x$  by  $y$ ;
2.  $\sigma(k) \in_n \text{Th}'$ ;
3. for some  $i, j < k$  we have  $\sigma(j) = \sigma(i) \rightarrow \sigma(k)$ .

$\text{Pr}_{\text{Th}'}(\psi)$  abbreviates the formula constructed in the above way.

**Definition 42.**  $\text{Con}(\text{Th}')$  denotes the sentence  $\neg \text{Pr}_{\text{Th}'}(\ulcorner 0 = 1 \urcorner)$ .

**Example 43 (PA).**  $\emptyset$  is the  $\Delta_0$  arithmetical theory; i.e. the empty set of sentences. Consequently,  $\text{Pr}_{\emptyset}(x)$  holds if  $x$  is provable in pure First Order Logic.  $I\Sigma_x$  is the arithmetical  $\Delta_1$  theory consisting of all sentences  $\text{Ind}(v, \phi)$ , for  $v$  a variable and  $\phi$  - a  $\Sigma_x$  formula, and all the axioms of  $Q$ . PA is the arithmetical  $\Delta_1$  theory consisting of all axioms of  $Q$  and sentences  $\text{Ind}(v, \phi)$ , for a variable  $v$  and an arithmetical formula  $\phi$  (we shall often write  $x \in \text{PA}$  instead of  $\text{PA}(x)$ ).

<sup>4</sup> The list of axioms of First-Order Logic is taken from [10].

We now turn to the development of model theory inside PA.

**Definition 44** (Th). A  $\Delta_n$  model  $M$  for arithmetic is a tuple  $\langle U_M, +_M, \cdot_M, a_0, a_1 \rangle$ , where

1.  $U_M, +_M, \cdot_M$  are  $\Delta_n$  sets,  $+_M \subseteq U_M^2$  and  $\cdot_M \subseteq U_M^2$
2.  $a_0, a_1$  are any numbers.

A full  $\Delta_n$  model  $\mathcal{M}$  is a triple  $\langle M, \text{val}_{\mathcal{M}}, \text{Sat}_{\mathcal{M}} \rangle$  such that

1.  $\mathcal{M}$  is a  $\Delta_n$  model.
2.  $\text{val}_{\mathcal{M}}$  is a  $\Delta_n$  function of evaluation of terms.
3.  $\text{Sat}_{\mathcal{M}}$  is a  $\Delta_n$  satisfaction relation for  $\mathcal{M}$ .

The precise definitions of  $\text{val}_{\mathcal{M}}$  and  $\text{Sat}_{\mathcal{M}}$  are as usual in model theory (details in the context of arithmetic being given in [19]). If  $\phi$  is any formula then function  $\alpha$  is an  $\mathcal{M}$  assignment for  $\phi$  if

1.  $\text{FV}(\phi) \subseteq \text{dom}(\alpha)$ ;
2.  $\text{im}(\alpha) \subseteq M$ .

**Convention 5.** We shall use the standard model-theoretic conventions when working with models in Th. In particular, working in Th, if  $\mathcal{M}$  is a full  $\Delta_n$  model, then for every formula  $\phi$  and  $\alpha$  an  $\mathcal{M}$  assignment for  $\phi$  we write

$$\mathcal{M} \models_{\mathcal{M}} \phi[\alpha]$$

for  $\langle \phi, \alpha \rangle \in_n \text{Sat}_{\mathcal{M}}$ . Moreover, if  $\phi$  is a sentence then  $\mathcal{M} \models_{\mathcal{M}} \phi$  stands for  $\mathcal{M} \models_{\mathcal{M}} \phi[\varepsilon]$ .

**Definition 45** (Th). Let  $\text{Th}'$  be any  $\Delta_n$  arithmetical theory. If  $\mathcal{M}$  is any full model, then we write  $\mathcal{M} \models_{\mathcal{M}} \text{Th}'$  if for every  $\phi \in \text{Th}'$ ,  $\mathcal{M} \models_{\mathcal{M}} \phi$ .

Proof of the following proposition proceeds by a straightforward induction on the lengths of proofs, carried out in PA.

**Proposition 46** (Th). Let  $\text{Th}'$  be any  $\Delta_n$  arithmetical theory and  $\mathcal{M}$  any full model such that  $\mathcal{M} \models \text{Th}'$ . Then for arbitrary formula  $\phi \in \mathcal{L}_{\text{Th}'}$  we have

$$\text{Pr}_{\text{Th}'}(\phi) \rightarrow \mathcal{M} \models_{\mathcal{M}} \phi$$

**Convention 6.** Let  $\mathcal{N} \models \text{Th}$  and  $\mathcal{N} \models$  "  $\mathcal{M}$  is a full arithmetic model ". Then  $\mathcal{M}$  canonically encodes a model in the standard set-theoretic sense, i.e. model  $\mathcal{M}'$  with universe

$$M' := \{x \in N \mid \mathcal{N} \models U_M(x)\}$$

and the interpretation of  $+$ ,  $\cdot$ ,  $0$ ,  $1$  defined analogously using  $+_M$ ,  $\cdot_M$ ,  $a_0$  and  $a_1$ . Moreover for every formula  $\phi$  of  $\mathcal{L}_{\text{PA}}$  and every  $\mathcal{M}$  assignment  $\alpha$  for  $\phi$  we have

$$\mathcal{M}' \models \phi[\alpha] \iff \mathcal{N} \models \left( \mathcal{M} \models_{\mathcal{M}} \ulcorner \phi \urcorner [\alpha] \right),$$

where  $\models$  denotes the standard set-theoretical satisfaction relation,  $\models_{\mathcal{M}}$  is an  $\mathcal{N}$ -definable satisfaction relation for  $\mathcal{M}$  and  $\alpha$  on the left-hand side is the (set-theoretical) function encoded by  $\alpha$ . Abusing the notation a little, we shall use the same letters for arithmetic models and their set-theoretic counterparts.

Let us note that if  $\phi(x)$  is a formula, such that  $\ulcorner \phi(x) \urcorner$  provably in PA defines a  $\Delta_n$  theory, then in every model  $\mathcal{M} \models \text{PA}$ , the set

$$\phi^{\mathcal{M}} \cap \mathbb{N} := \{n \in \mathbb{N} \mid \mathcal{M} \models \phi(n)\}$$

codes (via the translation from Theorem 13) a theory in the traditional sense of ZFC. However, if  $\phi$  is of high complexity in the sense of the  $\Sigma_n$ -hierarchy, then this theory might highly depend on the choice of  $\mathcal{M}$ . For example, if

$$\phi(x) := ((x = \ulcorner 0 = 0 \urcorner) \wedge \text{Con}(\text{PA})) \vee ((x = \ulcorner 0 = 1 \urcorner) \wedge \neg \text{Con}(\text{PA}))$$

there are models  $\mathcal{M}, \mathcal{M}'$  such that  $\phi^{\mathcal{M}} \cap \mathbb{N} = \{\ulcorner 0 = 0 \urcorner\}$  and  $\phi^{\mathcal{M}'} \cap \mathbb{N} = \{\ulcorner 0 = 1 \urcorner\}$ . Such a situation is impossible, if  $\ulcorner \phi(x) \urcorner$  is provably in PA a  $\Delta_1$  theory, since all  $\Delta_1$  sets of natural numbers are *strongly representable* in PA:

**Definition 47.** Let  $A \subseteq \mathbb{N}$ . We shall say that  $A$  is *strongly represented* in PA, if there exists a formula of  $\mathcal{L}_{\text{PA}}$   $\phi(x)$ , such that for all  $n \in \mathbb{N}$

$$n \in A \Rightarrow \text{PA} \vdash \phi(\underline{n}) \tag{2.1}$$

$$n \notin A \Rightarrow \text{PA} \vdash \neg \phi(\underline{n}) \tag{2.2}$$

The well-known fact in metamathematics of PA is that "being strongly represented in PA" is the same as "being  $\Delta_1$  definable":

**Theorem 48.** *The set  $A \subseteq \mathbb{N}$  is strongly represented if and only if for some  $\Delta_1(\text{PA})$  formula  $\phi(x)$  and for all  $n \in \mathbb{N}$ :*

$$n \in A \iff \mathbb{N} \models \phi(n)$$

For the proof of this theorem, see [24] or [36] (in Polish).

**Convention 7.** If  $\text{Th}'$  is provably in Th a  $\Delta_1$  arithmetical theory, then we shall associate with it the set of arithmetical sentences defined by  $\text{Th}'$  in the standard model  $\mathbb{N}$ . Abusing the notation, we shall denote this theory also by  $\text{Th}'$  and write, for example,

$$\text{Th}' \cup \{\text{Pr}_{\text{Th}'}(\ulcorner \phi \urcorner) \rightarrow \phi \mid \phi \in \text{Sent}_{\mathcal{L}_{\text{PA}}}\}$$

where first  $\text{Th}'$  denotes the set of arithmetical sentences and  $\text{Th}'$  written in  $\text{Pr}_{\text{Th}'}(x)$  denotes the chosen arithmetical formula. Basing on this convention, instead of writing that Th is a provably in PA,  $\Delta_1$  theory, we shall call it simply a  $\Delta_1$  theory.

The theorem below was proved in [19], Theorem 4.27 (Chapter 1), for the case of  $n = 1$  and  $\text{Th} = \text{PA}$  but the proof in greater generality is essentially the same.

**Theorem 49** (Arithmetised Completeness Theorem). *For every  $n \geq 1$ , Th proves that every consistent  $\Delta_n$  arithmetical theory has a full  $\Delta_{n+1}$  model.*

*Translations between logics*

In the course of our thesis, we shall compare theories of truth formulated for the arithmetised language extending  $\mathcal{L}_{\text{PA}}$  with new logical connectives and quantifiers. In the respective fragments, the notion of a *translation between logics* will play a key role.

**Definition 50.** A *logical signature*  $\sigma$  is a triple  $(q_\sigma, o_\sigma, \tau_\sigma)$  such that  $q_\sigma$  and  $o_\sigma$  are non-empty finite sets (thought of as the sets of *quantifiers* and *propositional connectives* respectively) and  $\tau_\sigma$  is a function with domain  $q_\sigma \cup o_\sigma$  and codomain  $\omega$ . The values of  $\tau$  are called *arities*. The arities determine how many formulae the given logical symbol takes. When we are given a logical signature  $\sigma$ , a set of non-logical constants  $\lambda$  (i.e. predicate symbols, function symbols and individual constants; we assume that symbols come with prescribed arities) and a set of variables  $\text{Var}$  then this triple is called a *language*. When we are given a language  $\mathcal{L}$  we can write the formula defining the set of formulae of  $\mathcal{L}$ ,  $\text{Form}_{\mathcal{L}}$ , in the following way: the notion of a term is as in any standard textbook (e.g. [10]) and now we define  $\text{Form}_{\mathcal{L}}$  as the least set satisfying:

1. if  $t_0, \dots, t_{n-1}$  are any terms and  $R$  is any predicate symbol of arity  $n$ , then  $R(t_0, \dots, t_{n-1}) \in \text{Form}_{\mathcal{L}}$ ;
2. if  $\phi_0, \dots, \phi_{n-1} \in \text{Form}_{\mathcal{L}}$  and  $O \in o_\sigma$  is such that  $\tau_\sigma(O) = n$ , then

$$O(\phi_0, \dots, \phi_{n-1}) \in \text{Form}_{\mathcal{L}}$$

3. if  $\phi_0, \dots, \phi_{n-1} \in \text{Form}_{\mathcal{L}}$ ,  $v \in \text{Var}$  and  $Q \in q_\sigma$  is such that  $\tau_\sigma(Q) = n$ , then

$$Q(v; \phi_0, \dots, \phi_{n-1}) \in \text{Form}_{\mathcal{L}}$$

Sometimes instead of  $\phi \in \text{Form}_{\mathcal{L}}$  we write  $\phi \in \mathcal{L}$  for short.

**Remark 51.** The above definition is not utterly general: for example, we do not allow for quantifiers binding more than one variable at the same time. This level of generality is enough for our purposes: we will never consider more exotic logics.

**Definition 52** (Translation between languages). Suppose  $\sigma = (q_\sigma, o_\sigma, \tau_\sigma)$  and  $\sigma' = (q_{\sigma'}, o_{\sigma'}, \tau_{\sigma'})$  are two logical signatures. Let  $\lambda$  and  $\text{Var}$  be two sets of extra-logical symbols and variables. Let  $\mathcal{L}_\sigma$  be the language over  $\sigma$ ,  $\lambda$  and  $\text{Var}$  and  $\mathcal{L}_{\sigma'}$  be the language over  $\sigma'$ ,  $\lambda$  and  $\text{Var}$ . Moreover let  $\mathbb{P} = \{p_i \mid i \in \omega\}$  be a set of fresh propositional constants (predicates of arity 0) that are neither elements of  $\mathcal{L}_\sigma$  nor elements of  $\mathcal{L}_{\sigma'}$ . Let  $\mathcal{L}_{\sigma'}^*$  be a language  $\mathcal{L}_{\sigma'}$  with additional constants from  $\mathbb{P}$ . Formulae of  $\mathcal{L}_{\sigma'}^*$  are to be thought of as propositional templates for formulae of  $\mathcal{L}_{\sigma'}$ . If

$$\phi, \psi_0, \dots, \psi_n$$

are any formulae of  $\mathcal{L}_{\sigma'}^*$ , then

$$\phi[\psi_0/p_0, \dots, \psi_n/p_n]$$

denotes the result of substituting  $\psi_k$  for  $p_k$  in  $\phi$  for every  $k \leq n$ . A *pre-translation* of  $\mathcal{L}_\sigma$  into  $\mathcal{L}_{\sigma'}$  is a function

$$\rho : (q_\sigma \times \text{Var}) \cup o_\sigma \rightarrow \text{Form}_{\mathcal{L}_{\sigma'}^*},$$

such that for every  $h \in (q_\sigma \times \text{Var}) \cup o_\sigma$  and every  $n$  if

1.  $h = \langle Q, v \rangle$  for some quantifier  $Q$  and variable  $v$  and  $\tau(Q) = n$  or
2.  $h \in o_\sigma$  and  $\tau(h) = n$ ,

then  $\rho(h)$  is a formula  $\phi \in \text{Form}_{\mathcal{L}_{\sigma'}}$  with exactly  $p_0, \dots, p_{n-1}$  as propositional constants. Function

$$* : \text{Form}_{\mathcal{L}_\sigma} \rightarrow \text{Form}_{\mathcal{L}_{\sigma'}}$$

is a *translation* of  $\mathcal{L}_\sigma$  into  $\mathcal{L}_{\sigma'}$  if and only if there exists a pre-translation  $\rho$  of  $\mathcal{L}_\sigma$  into  $\mathcal{L}_{\sigma'}$  such that the following conditions hold

1. for every predicate  $R \in \lambda$  of arity  $n$  and every terms  $t_0, \dots, t_{n-1}$

$$(R(t_0, \dots, t_{n-1}))^* = R(t_0, \dots, t_{n-1})$$

2. for every quantifier  $Q \in q_\sigma$  such that  $\tau(Q) = n$ , every formulae  $\phi_0, \dots, \phi_{n-1} \in \text{Form}_{\mathcal{L}_\sigma}$  and every variable  $v \in \text{Var}$  we have

$$(Q(v; \phi_0, \dots, \phi_{n-1}))^* = \rho(\langle Q, v \rangle)[\phi_0^*/p_0, \dots, \phi_{n-1}^*/p_{n-1}]$$

3. for every logical operator  $O \in o_\sigma$  such that  $\tau(O) = n$  and every sequence of formulae  $\phi_0, \dots, \phi_{n-1} \in \text{Form}_{\mathcal{L}_\sigma}$  we have

$$(O(\phi_0, \dots, \phi_{n-1}))^* = \rho(O)[\phi_0^*/p_0, \dots, \phi_{n-1}^*/p_{n-1}]$$

If  $*$  is a translation between  $\mathcal{L}_\sigma$  and  $\mathcal{L}_{\sigma'}$  and  $\mathcal{L}$  is an arbitrary language, we say that  $*$  is  $\mathcal{L}$ -conservative if for every formula  $\phi \in \text{Form}_{\mathcal{L}_\sigma} \cap \text{Form}_{\mathcal{L}}$  we have

$$\phi^* = \phi$$

Since the above definition refers solely to finite objects that admits recursive definitions, there are no obstacles in stating it in ZFC – Inf\* and, consequently, formalising it in PA.

### 2.0.3 Models of PA

**Convention 8.** If  $\mathcal{M}$  is a model for  $\mathcal{L}_P$ ,  $\phi(\bar{x})$  is an  $\mathcal{L}_P$  formula  $\bar{a}$  is a tuple of elements of  $\mathcal{M}$ , then we will often write  $\phi^{\mathcal{M}}(\bar{a})$  instead of  $\mathcal{M} \models \phi[\bar{a}]$ . Moreover,  $\phi^{\mathcal{M}}$  denotes the set of all tuples  $\bar{a}$  such that  $\phi^{\mathcal{M}}(\bar{a})$ .

**Convention 9 (Extensions).** Let  $\mathcal{M}, \mathcal{N}$  be two models of  $Q$  (possibly for a language  $\mathcal{L}_P$  extending  $\mathcal{L}_{\text{PA}}$ ) and let  $\mathcal{L} \subseteq \mathcal{L}_P$  be a language. We write

1.  $\mathcal{M} \subseteq_e \mathcal{N}$ , if  $\mathcal{M}$  is an end extension of  $\mathcal{M}$ .
2.  $\mathcal{M} \prec_{\mathcal{L}} \mathcal{N}$ , if  $\mathcal{M}$  is an  $\mathcal{L}$ -elementary submodel of  $\mathcal{N}$ .
3.  $\mathcal{M} \prec_{\mathcal{L},e} \mathcal{N}$ , if  $\mathcal{N}$  is an  $\mathcal{L}$ -elementary end extension of  $\mathcal{M}$ .

All the above notions are as defined in [27] and [24].

For the proof of the following proposition see [19], the first part of the proof of Theorem 2.40 (Chapter 3).

**Proposition 53.** *Let  $\mathcal{M} \models \text{Th}$  and let*

$$\mathcal{M} \models \mathcal{N} \text{ is a full } \Delta_n \text{ model of } Q$$

*Then  $\mathcal{M} \subseteq_e \mathcal{N}$ .*

**Convention 10.**

1. We will use the same symbol  $\omega$  to denote the least initial segment in every model of  $Q$ . We use  $a >^{\mathcal{M}} \omega$  to express the fact that for every  $n \in \omega$

$$\mathcal{M} \models a > n$$

i.e.  $a$  is a *nonstandard* element.

2. If  $\mathcal{M} \models \text{PA}$ , then  $\mathcal{L}_{\mathcal{M}}$  denotes the language of model  $\mathcal{M}$ , i.e. the language over arithmetical signature enriched with a constant for every element of  $\mathcal{M}$ .

For the proof of the next lemma see [24], Lemma 6.1.

**Lemma 54 (Overspill).** *Let  $\text{Th}$  be a fully inductive  $\mathcal{L}_P$  theory. Let  $\mathcal{M} \models \text{Th}$  and  $\phi(x)$  be a formula (possibly containing parameters from  $M$ ). Suppose that for every  $n \in \omega$*

$$\mathcal{M} \models \phi(n).$$

*Then for some  $a >^{\mathcal{M}} \omega$  we have*

$$\mathcal{M} \models \phi(a)$$

**Definition 55 (Class).** Let  $\mathcal{M} \models \text{PA}$ . A set  $A \subseteq M$  is a *class* if every  $M$ -finite fragment of  $A$  is coded in  $\mathcal{M}$ , i.e. for every  $c \in M$  there exists a  $d \in M$  such that

$$\{a \in A \mid a <^{\mathcal{M}} c\} = \{a \in M \mid a \in^{\mathcal{M}} d\}$$

We shall say that  $A \subset M^n$  is a *class* if the set

$$\{\langle a_0, \dots, a_{n-1} \rangle \in M \mid (a_0, \dots, a_{n-1}) \in A\}$$

is a class.

The proposition below can be found in [27] (Proposition 1.4.2). We sketch the proof for the reader's convenience

**Proposition 56.** *[Characterization of  $\Delta_0$ -induction] Let  $P$  be an  $n$ -ary predicate and  $\text{Th}$  be a theory in  $\mathcal{L}_P$  extending PA. The following are equivalent:*

1.  $\text{Th} \vdash I\Delta_0(\mathcal{L}_P)$

2. in arbitrary  $\mathcal{M} \models \text{Th}$ , (the extension of)  $P$  is a class.

*Sketch of the proof.* Without loss of generality assume that  $P$  is a unary predicate for otherwise define

$$P_1(y) := \exists x_0 < y \dots \exists x_{n-1} < y \ (y = \langle x_0, \dots, x_{n-1} \rangle \wedge \bigwedge_{i < n} P(x_i))$$

Then  $\text{Th} \vdash I\Delta_0(\mathcal{L}_P)$  iff  $\text{Th} \vdash I\Delta_0(\mathcal{L}_{P_1})$  and by definition, in every model  $\mathcal{M} \models \text{Th}$ ,  $P^\mathcal{M}$  is a class iff so is  $P_1^\mathcal{M}$ .

1.  $\Rightarrow$  2. Suppose  $\text{Th} \vdash I\Delta_0(\mathcal{L}_P)$  and fix an arbitrary  $\mathcal{M} \models \text{Th}$ , and an arbitrary  $c \in M$ . Let  $d$  be the set of all elements below  $c$  i.e. for all  $x \in M$

$$x \in^\mathcal{M} d \iff x <^\mathcal{M} c$$

Let us consider the following formula

$$\phi(x) := x < c + 1 \rightarrow \exists y < d + 1 \forall z < x \ (z \in y \equiv P(z))$$

$\phi(x)$  is of class  $\Delta_0$  and by a routine argument using induction on  $x$  we show

$$\forall x \phi(x)$$

proving 2.

2.  $\Rightarrow$  1. Let us fix an arbitrary  $\mathcal{M} \models \text{Th}$ , and an arbitrary formula  $\phi(x) \in \Delta_0$ . Since all quantifiers in  $\phi$  are bounded and  $P$  is a class, then for all  $a \in M$  there exists a  $c \in M$  such that for all  $d \leq^\mathcal{M} a$ ,

$$\mathcal{M} \models \phi(d) \iff \mathcal{M} \models \phi[y \in c/P(y)](d)$$

where  $\phi[y \in c/P(y)]$  is as explained in Definition 7 (the above property can be proved by induction on the complexity of  $\phi$ , using the fact that all quantifiers occurring in it are bounded). Suppose for some  $a \in M$

$$\mathcal{M} \models \phi(0) \wedge \neg\phi(a)$$

Let  $c$  code the set of elements satisfying  $P$  below  $a$ . We have

$$\mathcal{M} \models \phi[x \in c/P](0) \wedge \neg\phi[x \in c/P](a)$$

Hence, by the least number principle in  $\mathcal{M}$  (we use the fact, that  $\mathcal{M} \models \text{PA}$ ) there exists a  $d$  such that

$$\mathcal{M} \models \neg\phi[y \in c/P(y)](d) \wedge \forall x < d \phi[x \in c/P(y)](x)$$

In consequence  $\mathcal{M} \models \neg\phi(d) \wedge \forall x < d \phi(x)$ , which ends the proof.  $\square$

The next remark will play an important role in justifying the use of  $\Delta_0$  induction for formulae not in  $\Delta_0$  form.

**Remark 57.** Suppose  $\text{Th}$  is a  $\mathcal{L}_P$  theory proving  $I\Delta_0(\mathcal{L}_P)$  and extending PA. Suppose that  $\phi$  is an  $\mathcal{L}_P$  formula, such that (working in  $\text{Th}$ ) for some  $z$  we have

$$\phi(y) \equiv \phi[(P(\bar{x}) \wedge \langle \bar{x} \rangle < z) / P(\bar{x})](y)$$

i.e. only a bounded fragment of  $P$  is needed to determine the extension of  $\phi(y)$ . Then  $\text{Th}$  proves the induction axiom for  $\phi(y)$ . Indeed, working in  $\text{Th}$  with the same  $z$  as previously, by the above assumption and Proposition 56 there exists  $c$  such that

$$\forall \bar{x} \left( (P(\bar{x}) \wedge \langle \bar{x} \rangle < z) \equiv \langle \bar{x} \rangle \in c \right)$$

In particular, we have

$$\phi(y) \equiv \phi[\langle \bar{x} \rangle \in c / P(\bar{x})](y)$$

Hence,  $\phi(y)$  equivalent to an arithmetical formula (with parameters). Since  $\text{Th}$  extends PA, we are allowed to use the induction axiom for  $\phi[\langle \bar{x} \rangle \in c / P(\bar{x})](y)$  and, consequently, for  $\phi(y)$ .

**Definition 58** (Recursive Saturation). Let  $\mathcal{M} \models \text{PA}$ .

1. Let  $p(x) = \{\phi_i(x, \bar{a}) \mid i \in \omega\}$  be set of  $\mathcal{L}_{\mathcal{M}}$  formulae with free variable  $x$  and using finitely many parameters from  $\mathcal{M}$ .  $p(x)$  is a *type over  $\mathcal{M}$*  if for every  $n \in \omega$

$$\mathcal{M} \models \exists x \bigwedge_{i \leq n} \phi_i(x, \bar{a}).$$

The type  $p(x)$  is *recursive* if additionally the set  $\{\ulcorner \phi(x, \bar{y}) \urcorner \mid \phi(x, \bar{a}) \in p(x)\}$  is recursive.

2. The type  $p(x)$  is *realised in  $\mathcal{M}$*  if there exists  $a \in M$  such that for every  $\phi(x) \in p(x)$ ,  $\mathcal{M} \models \phi(a)$ .
3.  $\mathcal{M}$  is called *recursively saturated* if every recursive type over  $\mathcal{M}$  is realised in  $\mathcal{M}$ .

The next proposition is an immediate consequence of Löwenheim-Skolem Theorem (see e.g. [35]) and the fact that every countable model is an elementary submodel of a countable and recursively saturated model (see [24]). Its most important consequence is that for checking whether a sentence  $\phi$  is provable in PA, it is enough to examine whether it holds in every countable and recursively saturated model of PA.

**Proposition 59.** For every  $\mathcal{M} \models \text{PA}$  there exists a countable and recursively saturated model  $\mathcal{N}$  such that  $\mathcal{M}$  and  $\mathcal{N}$  are elementary equivalent.

The next definition introduces one of the most basic notions in investigating strength of truth theories. We will explain its significance in the next chapter.

**Definition 60** (Expandability). Let  $\mathcal{M} \models \text{PA}$  and let  $\text{Th}$  be a  $\mathcal{L}_P$  theory. Let  $n$  be the arity of  $P$ . We say that  $\mathcal{M}$  *expands* to a model of  $\text{Th}$  if and only if there exists  $A \subseteq M^n$  such that  $(\mathcal{M}, A) \models \text{Th}$ .

## 2.0.4 Arithmetical Reflection

Definitions below are to systematise the notation used to denote various extensions of arithmetical theories with reflection principles. Various such extensions have been studied in [1] and [11], for example (for a good introduction to the results presented in the second paper, see [16]). Also [8] gives a very good historical background to the investigations of reflection principles. The most important result of this section is Theorem 70, which is rather folklore to the discipline, but since it plays an important role in one of our arguments we reprove it for the reader's convenience.

**Definition 61** (Reflection over a theory). Let  $\text{Th}$  be any  $\Delta_1$  theory. We define  $\mathcal{UR}(\text{Th})$  to be the extension of  $\text{Th}$  with all sentences of the form

$$\forall t_0, \dots, t_n (\text{Pr}_{\text{Th}}(\ulcorner \phi(t_0, \dots, t_n) \urcorner) \rightarrow \phi(t_0^\circ, \dots, t_n^\circ))$$

for  $\phi(x_0, \dots, x_n) \in \text{Form}_{\mathcal{L}_{\text{PA}}}$

The proof of the Proposition below can be found in [19].

**Proposition 62** (Mostowski). *For every  $n$ ,  $\text{PA} \vdash \mathcal{UR}(I\Sigma_n)$ . In particular,  $\text{PA} \vdash \mathcal{UR}(\emptyset)$ .*

**Definition 63** (First reflective limit of a theory). Let  $\text{Th}$  be any  $\Delta_1$  theory. We define

$$\mathcal{UR}^0(\text{Th}) := \text{Th}$$

Suppose that the  $\Delta_1$  theory  $\mathcal{UR}^n(\text{Th})$  has been defined. Define

$$\mathcal{UR}^{n+1} := \mathcal{UR}(\mathcal{UR}^n(\text{Th}))$$

Let  $\mathcal{UR}^{n+1}$  be the natural  $\Delta_1$  formula strongly representing the above set of sentences. Finally the *first reflective limit of a theory* is the set

$$\bigcup_{n \in \omega} \mathcal{UR}^n(\text{Th})$$

We shall denote this set by  $\mathcal{UR}^\omega(\text{Th})$ .

Note that in the above definition we consider uniform reflection for *arithmetical* formulae only.

Shortly, we shall introduce one of the most basic theories of truth, which will simplify the reasoning in the proof Theorem 70. In fact, its use in this proof is the only reason for stating some of the previous definitions and propositions (like the above Arithmetised Completeness Theorem) for arbitrary fully inductive  $\mathcal{L}_P$  theories instead of simply PA:

**Definition 64** (Definition of UTB).  $\text{UTB}^-$  is the  $\mathcal{L}_T$  theory extending PA with all sentences of the form

$$\forall t_0, \dots, t_n (T(\ulcorner \phi(t_0, \dots, t_n) \urcorner) \equiv \phi(t_0^\circ, \dots, t_n^\circ))$$

for every formula  $\phi(x_0, \dots, x_n) \in \mathcal{L}_{\text{PA}}$ .  $\text{UTB}$  is the fully inductive extension of  $\text{UTB}^-$ .

Let us list the reasons why UTB is of interest to the current study:

1. it is fully inductive: hence, all the above definitions, stated usually in PA, make sense in UTB and all the above theorems are provable within it;
2. it proves no more arithmetical theorems (i.e. sentences of  $\mathcal{L}_{\text{PA}}$ ) than PA. More concretely, the following proposition (its proof can be found in [20]) holds:

**Proposition 65.** *For every sentence  $\phi \in \mathcal{L}_{\text{PA}}$ , if  $\text{UTB} \vdash \phi$ , then  $\text{PA} \vdash \phi$ .*

3. in every model  $\mathcal{M}$  of UTB the extension of  $T$  contains the elementary diagram of  $\mathcal{M}$ . More precisely, if  $(\mathcal{M}, T) \models \text{UTB}$ ,  $\bar{a} \in M$  and  $\phi$  is an  $\mathcal{L}_{\text{PA}}$  formula such that  $\mathcal{M} \models \phi[\bar{a}]$ , then  $\ulcorner \phi(\bar{a}) \urcorner \in T$ .
4. it imposes recursive saturation; more precisely, we have the following theorem:

**Theorem 66.** *If  $\mathcal{M} \models \text{UTB}$ , then  $\mathcal{M}$  is recursively saturated.*

For the details see [24], Proposition 15.4<sup>5</sup>

The following proposition is an easy consequence of the second point of the above list (in fact, it is an easy consequence of compactness and the existence of partial truth predicates, as in Definition 33; for the details of its proof, see [24], Proposition 15.1<sup>6</sup>

**Proposition 67.** *For every  $\mathcal{M} \models \text{PA}$ , there exists  $\mathcal{N} \models \text{UTB}$  such that  $\mathcal{M} \prec_{\mathcal{L}_{\text{PA}}} \mathcal{N}$ .*

In particular, we get the following refinement of Completeness Theorem:

**Corollary 68.** *For every sentence  $\phi \in \mathcal{L}_{\text{PA}}$ ,  $\text{PA} \vdash \phi$  if and only if for every  $\mathcal{M} \models \text{UTB}$ ,  $\mathcal{M} \models \phi$ .*

The next theorem ends our preliminary section: it will be of crucial importance to Section 5.3 and will be of no use before that. It will be convenient to introduce one more definition (actually it is a scheme of definition, for every  $n, k$ ).

**Definition 69.** Let  $\text{Th}'$  be any  $\Delta_1$  theory.  $\text{Tr}_{\text{Th}'} \upharpoonright_k$  is the Th provable arithmetical theory defined

$$\text{Tr}_{\text{Th}'} \upharpoonright_k (x) := x \in_n \text{Th}' \vee \text{Tr}_k(x).$$

Moreover we define

$$\text{SCon}(\text{Th}') := \{\text{Con}(\text{Tr}_{\text{Th}'} \upharpoonright_n) \mid n \in \omega\}$$

**Theorem 70.** *Let  $\mathcal{M} \models \text{UTB}$  and suppose that  $\text{Th}$  is an arithmetically definable arithmetical  $\Delta_n$  theory. If  $\mathcal{M} \models \text{SCon}(\text{Th})$  then there exists full arithmetical model  $\mathcal{N}$  such that*

1.  $\mathcal{M} \prec_{\mathcal{L}_{\text{PA}}} \mathcal{N}$  and

<sup>5</sup> Kaye phrases this theorem in terms of *partial nonstandard inductive satisfaction classes*. For the proof of precisely this formulation, consult [33]. It can be easily seen that having a partial inductive satisfaction class and expanding to a model of UTB are equivalent conditions.

<sup>6</sup> As previously this theorem is phrased in terms of partial nonstandard inductive satisfaction classes

2.  $\mathcal{M} \models (\mathcal{N} \models_{\mathcal{N}} \text{Th})$ .

*Proof.* Let us fix  $\mathcal{M} \models \text{UTB}$  and suppose that for some arithmetically definable  $\Delta_n$  arithmetical theory  $\text{Th}$ , for each each  $k$ ,  $\mathcal{M} \models \text{Con}(\text{Tr}_{\text{Th}} \upharpoonright_k)$ . Let us define

$$\text{Th}'(v_0, w) := (v_0 \in_n \text{Th}) \vee (\text{Compl}(v_0) \leq w \wedge T(v_0))$$

Then for each  $c \in M$

$$\mathcal{M} \models \text{"Th}'(v_0, c) \text{ is a } \Delta_{n+1} \text{ } \mathcal{L}_T\text{-theory"}$$

Since for every  $n \in \omega$ ,

$$\mathcal{M} \models \forall \phi (\text{Compl}(\phi) \leq n \rightarrow (T(\phi) \equiv \text{Tr}_n(\phi))),$$

(the above sentence is provable in UTB by the external induction on  $n$ ), then for every  $n$  in  $\omega$  we have

$$\mathcal{M} \models \text{Con}(\text{Th}'(v_0, \underline{n}))$$

( $\text{Th}'(v_0, n)$  is a formula with one free variable  $v_0$  and numeral denoting  $n$ ). By overspill (Lemma 54) there is a  $c >^{\mathcal{M}} \omega$  such that

$$\mathcal{M} \models \text{Con}(\text{Th}'(v_0, \underline{c}))$$

Let  $\mathcal{N}$  be the full model of  $\text{Th}'(v_0, \underline{c})$ . Then

1.  $\mathcal{M} \models (\mathcal{N} \models_{\mathcal{N}} \text{Th})$
2.  $\mathcal{M} \prec_{\mathcal{L}_{\text{PA}}} \mathcal{N}$

The first point is obvious, since  $\mathcal{M} \models \text{Th} \subseteq_{n+1} \text{Th}'(v_0, \underline{n})$  (by the definition of  $\text{Th}(v_0, c)$ ). The second one holds since the conjunction of axioms of  $Q$ , let us denote it with  $\phi_Q$ , is a  $\Pi_1(\text{PA})$  sentence true in  $\mathcal{M}$ . Hence,  $\mathcal{M} \models T(\ulcorner \phi_Q \urcorner)$  and consequently

$$\mathcal{M} \models \ulcorner \phi_Q \urcorner \in_{n+1} \text{Th}'(v_0, c).$$

Hence, also  $\mathcal{N} \models \phi_Q$  (see Convention 6). Hence, by Proposition 53  $\mathcal{M} \subseteq \mathcal{N}$ . Moreover if  $\phi(x_0, \dots, x_n)$  is any  $\Sigma_n$  formula of complexity  $k$  and  $a_0, \dots, a_n$  are such that  $\mathcal{M} \models \phi[a_0, \dots, a_n]$ , then

$$\mathcal{M} \models \text{Compl}(\ulcorner \phi(\underline{a}) \urcorner) \leq k \wedge \Sigma_n(\ulcorner \phi(\underline{a}) \urcorner) \wedge T(\ulcorner \phi(\underline{a}) \urcorner)$$

where  $\ulcorner \phi(\underline{a}) \urcorner$  is a short for  $\ulcorner \phi(\underline{a}_0, \dots, \underline{a}_n) \urcorner$ . Consequently,  $\ulcorner \phi(\underline{a}) \urcorner \in_{n+1}^{\mathcal{M}} \text{Th}'(v_0, c)$ , and  $\mathcal{N} \models \phi(a_0, \dots, a_n)$ .  $\square$

The next proposition provides a link between Uniform Reflection and Strong Consistency.

**Lemma 71.** *For every  $\Delta_1$  theory  $\text{Th}$  extending  $I\Sigma_1$ , for every  $n, k, m$  and every formula  $\phi(x_0, \dots, x_m)$  in  $\mathcal{L}_{\text{PA}}$  we have*

$$\mathcal{UR}(\text{Th}) \vdash \forall x_0, \dots, x_m (\text{Pr}_{\text{Tr}_{\text{Th}} \upharpoonright_k}(\ulcorner \phi(\underline{x}_0, \dots, \underline{x}_m) \urcorner) \rightarrow \phi(x_0, \dots, x_m))$$

*Proof.* Let us fix  $\text{Th}$ ,  $n, m, k$  and a formula  $\phi(x_0, \dots, x_m) \in \mathcal{L}_{\text{PA}}$ . Working in  $\mathcal{UR}(\text{Th})$  assume

$$\text{Pr}_{\text{Tr}_{\text{Th}} \upharpoonright k} (\ulcorner \phi(\underline{x}_0, \dots, \underline{x}_m) \urcorner).$$

Hence, there are sentences  $\psi_0, \dots, \psi_a$  of complexity at most  $k$  such that

$$\text{Pr}_{\text{Th}} \left( \left( \bigwedge_{i \leq a} \psi_i \right) \rightarrow \phi(\underline{x}_0, \dots, \underline{x}_m) \right).$$

Since  $\Sigma_k(\emptyset)$  classes are closed under taking conjunctions, we have there exists a sentence  $\theta$  of complexity  $\Sigma_k$  such that

$$\text{Pr}_{\emptyset} \left( \left( \bigwedge_{i \leq a} \psi_i \right) \equiv \theta \right).$$

Hence, since  $\text{Th}$  extends  $I\Sigma_1$ , by Proposition 34, we have

$$\text{Pr}_{\text{Th}} (\text{Tr}_{\Sigma_k}(\theta) \rightarrow \phi(\underline{x}_0, \dots, \underline{x}_m)).$$

By the uniform reflection for  $\text{Th}$  and the formula

$$\text{Tr}_{\Sigma_k}(\theta) \rightarrow \phi(x_0, \dots, x_m)$$

we obtain

$$\text{Tr}_{\Sigma_k}(\theta) \rightarrow \phi(x_0, \dots, x_m),$$

which ends the proof, since  $\text{Tr}_{\Sigma_k}(\theta)$  holds.  $\square$

**Proposition 72.** *For every  $\Delta_1$  theory  $\text{Th}$  extending  $I\Sigma_1$ , the following theories are deductively equivalent*

1.  $\text{PA} + \mathcal{UR}(\text{Th})$
2.  $\text{PA} + \text{SCon}(\text{Th})$

Moreover if  $\text{Th}$  is a  $\Delta_1$  theory which is contained in  $I\Sigma_n$  for some  $n$ , then

$$\text{PA} \vdash \text{SCon}(\text{Th})$$

*Proof.* Let us start with the first part of the thesis.

1.  $\Rightarrow$  2.

It follows immediately by Lemma 71.

2.  $\Rightarrow$  1.

Let us fix arbitrary  $\mathcal{M} \models \text{PA} + \text{SCon}(\text{Th})$ . If for some  $\phi(x_0, \dots, x_n)$  and some arithmetised closed terms  $t_0, \dots, t_n \in M$  we had

$$\mathcal{M} \models \text{Pr}_{\text{Th}}(\phi(t_0, \dots, t_n)) \wedge \neg \phi(t_0^\circ, \dots, t_n^\circ)$$

then, by 36, we would have (for  $n = \text{Compl}(\phi)$ )

$$\mathcal{M} \models \text{Pr}_{\text{Th}}(\phi) \wedge \text{Tr}_{n+1}(\neg \phi)$$

Then we would have  $\mathcal{M} \models \text{Pr}_{\text{Tr}_{\text{Th}} \upharpoonright_{n+1}}(0 = 1)$ , contradicting  $\mathcal{M} \models \text{SCon}(\text{Th})$ .

The moreover part follows, since obviously it is sufficient to show that for each  $n \geq 1$

$$\text{PA} \vdash \text{SCon}(I\Sigma_n).$$

But for such theories this is a consequence of Theorem 62 and the first part of our proof.  $\square$

### 3. AXIOMATIC THEORIES OF TRUTH

#### 3.1 Definition and Motivations

On the formal side, in the field of Axiomatic Theories of Truth, we aim at determining properties of *axiomatic* theories built in the following way: we start with a base theory  $B$ , expressive enough to formalise syntax and enrich its language with a fresh unary predicate denoted  $T(x)$ . We then add axioms governing the newly added predicate. The minimal requirement for the resulting theory to be called *an axiomatic theory of truth over  $B$*  is to prove all sentences of the form

$$T(\ulcorner \phi \urcorner) \equiv \phi \quad (\text{T-scheme})$$

for every sentence  $\phi$  from the language of the base theory  $B$ . In the above  $\ulcorner \phi \urcorner$  denotes the canonical name of sentence  $\phi$ , according to the chosen way of formalising syntax in  $B$  (for  $B = \text{PA}$  it was introduced in the previous section).

**Example 73.** For  $B = \text{PA}$ , UTB, as defined in Definition 64, is an axiomatic theory of truth over PA.

Let us motivate the above sketch of definition: the base theory  $B$  is to serve as a model for all of our knowledge about facts that do not engage the notion of truth. Any axiomatic theory of truth over  $B$  models, then, what happens when this knowledge is supplemented with some statements expressing properties of the notion of truth? Let us observe that if  $B$  is consistent, then the predicate  $T$  added to  $B$  and satisfying T-scheme cannot be provably in  $B$  equivalent to any formula of  $\mathcal{L}_B$ . This is the content of the celebrated Tarski's Theorem (we focus on  $\mathcal{L}_P$ -theories but its applicability is far more general):

**Theorem 74** (Tarski; [34]). *Let  $B$  be any consistent  $\mathcal{L}_P$ -theory extending PA. There is no formula  $\theta(x)$  in  $\mathcal{L}_P$  such that for all  $\phi \in \mathcal{L}_P$*

$$B \vdash \theta(\ulcorner \phi \urcorner) \equiv \phi$$

where  $\ulcorner \cdot \urcorner$  is as explained in previous chapter.

In particular, extending  $B$ , somehow, is necessary for introducing the notion of truth for  $\mathcal{L}_B$  and we are interested in "what happens" when the truth predicate is introduced in the above-described way.

Such a way of formally introducing the notion of truth for a language is certainly not the only possibility. From the very beginning, it is worth to contrast our axiomatic approach with two others, both stemming from the work of Alfred Tarski: the study originally pursued by

Tarski himself, that initiated the field of Model Theory, and the "disentangled syntax approach", studied in [21] and [32]. Two issues are characteristic for the original Tarskian approach<sup>1</sup>:

1. The notion of truth for a theory  $B$  is introduced via definition which uses only non-semantical notions<sup>2</sup>;
2. The definition of truth is given in a metatheory, which properly extends the base theory  $B$ .

In fact, the motivation behind the second point is purely logical, assuming one wants to realise the first: by Tarski's Theorem 74, the truth definition cannot be given in the base theory, so it is necessary to enrich the language of  $B$ , somehow. That the first one was really what Tarski aimed at is amply illustrated by the following quotation:

It is desirable for the meta-language not to contain any undefined terms except such as are involved explicitly or implicitly in the remarks above, i.e.: terms of the object-language; terms referring to the form of the expressions of the object-language, and used in building names for these expressions; and terms of logic. In particular, we desire semantic terms (referring to the object-language) to be introduced into the meta-language only by definition. For, if this postulate is satisfied, the definition of truth, or of any other semantic concept, will fulfill what we intuitively expect from every definition; that is, it will explain the meaning of the term being defined in terms whose meaning appears to be completely clear and unequivocal. And, moreover, we have then a kind of guarantee that the use of semantic concepts will not involve us in any contradictions. ([44], 350-351)

Since the notion of truth should be analysable in terms of non-semantical notions, one has to introduce new primitive notions to the language of  $B$  (i.e. it is not sufficient to introduce new axioms in  $\mathcal{L}_B$ ), using Tarski's term: to enrich  $B$  essentially. For this reason, Tarski's approach cannot be used to model the situation we would like to study: if we have to enrich  $B$  with new notions of extra-semantical origins, then  $B$  could not encompass all of our extra-semantical knowledge. That is why, contrary to Tarski, we do not require that the truth predicate for the language of the base theory be *definable* in metatheory, in terms of more primitive notions, but it is a new *primitive* notion explicitly added to the language. To illustrate this distinction: Tarski showed (*inter alia*) that ZFC can define a truth predicate for PA, for example by putting

$$T_{\text{ZFC}}(\phi) := \mathbb{N} \models \phi \quad (\text{Tarski's Definition})$$

where  $\models$  denotes the satisfaction relation and  $\mathbb{N}$ —the standard model of arithmetic. All the notions on the right-hand side have well defined mathematical meaning, hence, the notion of

<sup>1</sup> In what follows, we are basing on [44]. What we call *theories*, Tarski would rather call *languages*. The distinction is purely terminological since for Tarski, any formal language always comes with specified axioms and rules of inference.

<sup>2</sup> Tarski's definition of truth uses the notion of *satisfaction*, but the latter is defined in terms of non-semantical notions.

arithmetical truth can be seen as *reduced* to the notion of a set (as axiomatised by ZFC<sup>3</sup>). For so defined  $T_{\text{ZFC}}$ , for every  $\phi$  belonging to the arithmetical language, ZFC proves

$$T_{\text{ZFC}}(\phi) \equiv \phi^* \quad (\text{ZFC-T-scheme})$$

where  $\phi^*$  is the standard translation of arithmetical sentence  $\phi$  to the set-theoretical metalanguage. In a sense, the approach undertaken in Axiomatic Theories of Truth is much more naïve: we assume that we have a notion of truth, which *a priori* need not have any mathematical content except for what we explicitly stipulate by adding some axioms. Pursuing this strategy, we might not succeed in laying mathematical foundations for the notion of truth by reducing it to the notion of purely mathematical provenance, but we can study the notion of truth for  $B$  without introducing objects or relations from the outside of  $B$ .

The second difference between our approach and the Tarskian model is where the syntactical notions should be formalised. What is interesting is that in [44], Tarski noticed that the metatheory of  $B$  need not be stronger than  $B$  itself,<sup>4</sup> if we stipulate that the truth predicate (or the satisfaction predicate) is introduced axiomatically:

Thus we see that the condition of "essential richness" is necessary for the possibility of a satisfactory definition of truth in the meta-language. If we want to develop the theory of truth in a meta-language which does not satisfy this condition, we must give up the idea of defining truth with the exclusive help of those terms which were indicated above (in Section 8). We have then to include the term "true," or some other semantic term, in the list of undefined terms of the meta-language, and to express fundamental properties of the notion of truth in a series of axioms. There is nothing essentially wrong in such an axiomatic procedure, and it may prove useful for various purposes.([44], p. 352)

When the notion of truth is introduced axiomatically, the question whether syntax of  $\mathcal{L}_B$  should be formalised inside of  $B$  or in a distinct metatheory for  $B$  is not addressed by Tarski (in [44]). However, in [21] and [32], Richard Heck (first paper), Graham Leigh and Carlo Nicolai (second paper) claim that mathematical and syntactical parts of the theory of truth for  $B$  should be kept separate. The main motivation for doing this is to isolate *purely* metatheoretical reasonings and to separate them from reasonings in the object theory, allowing the truth predicate only in the *former*. To illustrate the distinction, let us consider the following truth theory (we allow ourselves for informal presentation): we take ZFC as our base theory and extend its language with a binary predicate  $S(x, y)$  with the intended reading "sequence  $y$  satisfies formula  $x$ ", allowing it in the axioms schemata of comprehension and replacement. We also add the UTB-like scheme

$$\forall x_0, \dots, x_n (S(\phi, \langle x_0, \dots, x_n \rangle) \equiv \phi(x_0, \dots, x_n))$$

<sup>3</sup> In [44], Tarski admitted that metatheory should contain logic "in a broad sense"; i.e. including "the mathematical theory of sets". Under so robust understanding of logic, Tarski's truth definition can indeed be seen as given purely in logical terms. However, Tarski also stated that he preferred to treat logic more narrowly. See [44], footnote 12, p. 371

<sup>4</sup> In fact, it can be *much weaker* as all we demand is that some basic syntactical operations be expressible in it.

for all formulae  $\phi$  in the language of set-theory (where  $\langle x_0, \dots, x_n \rangle$  denotes the sequence of  $n$  elements having  $x_i$  as its  $i$ -th element). The prominent example of a metatheoretical reasoning involving the truth predicate is, for example, induction on the length of proofs used in demonstrating

$$\forall \phi \in \text{Sent}_{\text{ZFC}} (\text{Pr}_{\text{ZFC}}(\phi) \rightarrow S(\phi, \emptyset))$$

using the assumptions  $\forall \phi ((\text{Pr}_{\text{ZFC}}(0, \phi) \rightarrow S(\phi, \emptyset))$  and

$$\forall \phi \left( \forall n \in \omega \left( (\text{Pr}_{\text{ZFC}}(n, \phi) \rightarrow S(\phi, \emptyset)) \rightarrow (\text{Pr}_{\text{ZFC}}(n+1, \phi) \rightarrow S(\phi, \emptyset)) \right) \right),$$

where  $\text{Pr}_{\text{ZFC}}(x, \phi)$  formalises the relation "there exists a proof of sentence  $\phi$  from the axioms of ZFC in at most  $x$  steps". In such a case, the reasoning "... is carried out on syntactic objects" ([32]); i.e. derivations within a formal system. We reason in the object theory using the satisfaction predicate, when e.g. for a formula  $\phi$  in the sense of the object theory we would like to demonstrate

$$\forall x S(\phi, \langle x \rangle)$$

( $\langle x \rangle$  denotes the one-elementary sequence) via  $\in$ -induction, using the assumption

$$S(\phi, \langle \emptyset \rangle) \wedge \forall y \left( (\forall x \in y S(\phi, \langle x \rangle)) \rightarrow S(\phi, \langle y \rangle) \right)$$

For such an argument to be valid, the quantifiers really have to range over all the objects from the domain of our object theory, and not only over those of purely syntactical origins (obviously if the two cannot be identified, as in the example above).

In [32], metatheory is disentangled from the object theory by introducing two additional sorts of variables (i.e. new with respect to sorts of variables used by the object theory): one whose intended interpretation is the set of *expressions of the object language* and which can axiomatise relevant syntactical operations and the second for *sequences* (used as assignments). Then the syntax of the object theory is given in terms of objects of the first sort (by adding some appropriate axioms; in [32] the theory of expressions is simply PA) and a satisfaction predicate is characterized axiomatically as a relation between expressions and sequences of objects from the object domain of the object theory. In such a way, not only the truth predicate is "disentangled" from the base theory, which is a characteristic feature of our approach, but the whole metatheory is "disentangled" from it.

Let us now comment on the relation between our approach and that described above. In fact, what we have just seen is really a formalisation of Tarski's suggestion given in the second quotation above<sup>5</sup>. Such an approach has an obvious advantage of making it possible to study the notion of truth for theories that are too weak to encode syntax (such as the theory of fields<sup>6</sup>). Moreover, it provides a uniform way of building the metatheory for arbitrary theory Th and since the notion of truth is treated axiomatically, we might vary its axioms while keeping the syntactical part of the metatheory fixed, thus studying the notion of truth to some extent independently of the theory of syntax. Having said that, let us indicate why it is not the route

<sup>5</sup> This inspiration is explicit in both [21] and [32].

<sup>6</sup> Which, although undecidable, is not essentially undecidable, hence, cannot interpret  $Q$ .

we would like to take, indicating however that it is not because we find the "disentangled syntax" approach wrong in what it is meant to model. The main reason is really in the big picture we would like to investigate: we do not want to treat  $B$  only as a theory we reason *about*, but as a theory we reason *in*, when we do not use any semantical notions.  $B$  is to be treated as a model of our non-semantical knowledge and that is why we want to formalise syntax using the conceptual resources provided by  $B$  and not from the *outside* of  $B$ . Let us stress that in the "disentangled syntax" approach, we need not assume that the metatheory is "essentially richer" than the object theory. But still, it is based on the assumption that syntactical objects we use to characterise the truth for  $B$  do not belong to the object domain of  $B$ . As was suggested above, we do not want to treat all the metatheory as distinct from the base theory. The unique piece of our knowledge that comes to  $B$  from the outside is the knowledge about the notion of truth (or more general: semantics) for the language of  $B$ . The last crucial point is that when the truth (or satisfaction) predicate is added, we want it to characterise the notion of truth for the whole language of  $B$ ; i.e. for the language in which all non-semantical facts can be expressed. This is not the case in Tarski's and in the "disentangled syntax" approach, since there the notion of truth is introduced only for the object theory, leaving the notion of truth for the syntactical part of the metatheory undefined.

Although such a path is not exploited in our dissertation, let us briefly mention that our axiomatic approach leaves open the possibility of studying properties of the *self-applicable* notion of truth. It is a central idea of Tarski's truth definition that the truth predicate is defined for the language that does not contain it. As was convincingly argued by Kripke (in [30]), there are good reasons for developing theories of truth in which this restriction is lifted. Taking the axiomatic approach, the first step to realise this idea is trivial: simply add appropriate axioms to base theory  $B$ . Now we should check with how many axioms we can extend the base theory, not to render the resulting theory inconsistent. It showed up that many different axiomatisations of self-referential truth are possible (such as KF, WKF, PKF, FS, VF to name a few<sup>7</sup>, each highlighting an important trait of the notion of truth<sup>8</sup>

What are the questions we would like to answer? Our goal is to examine how the properties of a theory of truth change depending on the choice of truth-theoretical axioms. Consider, for example, the minimal theory of truth over PA:

**Definition 75 (TB).**  $TB^-$  is an  $\mathcal{L}_T$  theory extending PA with all axioms of the form

$$T(\ulcorner \phi \urcorner) \equiv \phi$$

where  $\phi$  is an arbitrary arithmetical sentence. TB is the fully inductive extension of  $TB^-$ .

Now we might ask whether adding induction to  $TB^-$  or considering  $UTB^-$  (Definition 64) instead of  $TB^-$  changes some properties of theories we investigate. Those properties, in turn,

<sup>7</sup> For the definition consult [20](KF, PKF, FS), [17] (WKF) and [3] (VF).

<sup>8</sup> For example, FS and PKF are "symmetrical", by which we mean that the *outer logic* (the logic in which respective theories are formulated) and *inner logic* (the logic for which the semantics given by the truth predicate is sound) coincide. This is not true of every theory of self-applicable truth: e.g. in KF both logics diverge, the inner being the three valued Strong Kleene Logic, the outer being classical. It is an interesting discovery that KF is far stronger than the two "symmetrical" systems.

serve as explications of certain other traits of a notion of truth. Let us give some examples of which properties of axiomatic theories might be considered, together with explanation of their philosophical meaning (in the following, Th is an arbitrary theory extending  $B$ ):

1. We might ask whether Th proves new (i.e. unprovable in  $B$ ) sentences of  $\mathcal{L}_B$ ;
2. We might ask whether Th facilitates proofs of theorems of  $B$ ; i.e. whether it proves some theorems of  $B$  much *faster* than  $B$  itself (this notion being fully explained originally in [38] and in this context in [13]);
3. We might ask which theories are relatively interpretable in Th (this notion being defined in [25] and [21]);
4. We might ask which models of  $B$  are expandable to models of Th.

We shall elaborate further on points 1, 3 and 4 later on in this chapter (see Section 3.2), since the three are actually our candidates for explicating the notion of "strength" of the notion of truth. Point 2 was introduced by Martin Fischer (in [13]) as an explication of the notion of "usefulness" of the notion of truth. In consequence, we may investigate how *strength* or the *usefulness* of the notion of truth is related to its various properties (represented by the chosen axioms for  $T$ ). For example: is classical compositionality a *useful* property of truth? Is classically compositional notion of truth *stronger*, than a non-classically compositional one? Under the taken approach, we might study how varying the axioms for  $T$  influences answers to any of the above questions. Furthermore, we can investigate which axioms for the truth predicate are responsible for proving new sentences in the language of  $B$ . Tarski Theorem states that the *full truth theory* over  $B$ , being inconsistent, proves a lot new sentences of  $\mathcal{L}_B$  (obviously if only  $B$  is consistent). Moreover, we know that if  $B$  is recursively axiomatisable, then there are sentences undecided by  $B$ . This is the content of the famous Gödel's Theorems:

**Theorem 76** (Gödel–Rosser's First Incompleteness Theorem; [34], [24], [19]). *If  $B$  is a consistent recursively axiomatisable theory extending  $Q$ , then there is a sentence  $\phi \in \mathcal{L}_B$  such that*

$$B \not\vdash \phi \text{ and } B \not\vdash \neg\phi$$

In the theorem below,  $\text{Con}(B)$  is the natural consistency statement for  $B$  as introduced in the previous chapter.

**Theorem 77** (Second Gödel's Theorem; [34], [19]). *If  $B$  is a consistent recursively axiomatisable theory extending  $Q$ , then  $B \not\vdash \text{Con}(B)$ .*

Now we might ask which axioms (*which truth properties*) are responsible for deciding undecidable sentences of the base theory. For many logicians, this view being clearly presented in [45], proving the consistency of  $B$  or reconstructing the metatheoretical reasoning that leads us to accept the Gödel sentence for  $B$ , is one of the most important roles of the notion of truth.

To better explain the above picture, let us draw a parallel between this programme and the motivations standing behind a plethora of research in Reverse Mathematics (for the exposition

of its most important results see [41]) and in fact, the Hilbert's Programme, since the former aims to "partially realise" the latter (this "partial realisation" is the main topic of [40]). The base theory is, in this case,  $I\Sigma_1$  being a "finitistically acceptable" part of mathematics, proving the same  $\Pi_1$  sentences<sup>9</sup> as *Primitive Recursive Arithmetic (PRA)*. The latter is usually taken (after the argumentation given by Tait in [43]) to be the theory capturing finitistic notions and reasonings. To this base theory we add axioms expressing properties of infinite sets of natural numbers, such as *Weak König's Lemma*, *Ramsey's Theorem* or *comprehension* for restricted class of formulae and then study whether those principles have *non-finitistic*  $\Pi_1$  consequences<sup>10</sup> Similarly to our case, this line of research aims at establishing what happens when principles concerning infinite totalities are added to purely finitistic base theory. We study what happens when principles concerning the truth are added to a base theory, which cannot represent this notion. We will go back to this comparison when explaining the notion of *strength*, which interests us most.

In our investigations we shall restrict our attention to one particular base theory  $B$  and study (a particular kind of) axiomatic theories of truth over Peano Arithmetic PA, as defined in Definition 5. There are at least four reasons to motivate this choice:

#### 1. Importance

PA is one of the most important theories in the Foundations of Mathematics being the natural first-order variant of Dedekind axiomatisation (that this is indeed the natural variant is sometimes called the *Isaacson's thesis*, see [8] for a discussion). Moreover, it can be treated as *the theory of hereditarily finite sets* in disguise, being bi-interpretable with a very natural set theory of hereditarily finite sets (see Subsection 2.0.2).

#### 2. Possibility of dealing with compositional truth

In contrast to e.g. ZFC, PA proves that every object from the domain can be *named* (by a numeral, for example). This feature makes it possible to define compositional axioms for  $T$  in the form

$$T(\exists v\phi) \equiv \exists xT(\phi(x))$$

whereas in ZFC such an axiom would yield unwanted consequences. In particular, to verify whether an existential sentence  $\exists x\phi$  is true, it would be sufficient to check whether an object *which can be named* satisfies  $\phi$ .

#### 3. Convenience

PA is a well-studied and fairly well understood axiomatic theory with nicely developed metamathematics (see e.g. [19]) and model theory (see e.g. [24], [27]).

#### 4. Not-that-much-loss-of-generality

Usually, it can be easily seen how results obtained for PA can be transferred to other base theories. This is, however, more heuristic than a theorem and the case of each axiomatic truth theory has to be considered separately.

<sup>9</sup> In fact, the same  $\Pi_2$  sentences, which means that the provably total in  $I\Sigma_1$  are exactly *primitive recursive functions*.

<sup>10</sup> The recent breakthrough in the field, obtained jointly by Ludovic Patey and Keita Yokoyama, shows that Ramsey Theorem for pairs and two colours is *finitistically reducible*.

In summary, we think that from a methodological point of view the choice of PA as a starting point is very natural: moreover, it is a well established line of research. For example, [20] is wholly devoted to axiomatic theories of truth extending PA. Obviously, it does not mean that there are no results concerning axiomatic truth or satisfaction theories obtained in greater generality. For example, [9] studies the theory of Full Satisfaction Class over every  $B$  strong enough for representing syntax (this latter notion being formalised in terms of interpretability of some designated theory; a similar perspective is taken up in [31]) and [18] investigates the strength of various theories of satisfaction over some set theories. Now the definition of axiomatic theory of truth as used in this dissertation can be given:

**Definition 78.** Let  $\mathcal{L}_T$  be the language  $\mathcal{L}_{PA} \cup \{T\}$  where  $T$  is a unary predicate symbol. An *axiomatic theory of truth* is any  $\mathcal{L}_T$  theory extending  $TB^-$ .

### 3.2 Strength of Axiomatic Theories of Truth

As already signalled in the last section, the main focus of this dissertation is the *strength* of axiomatic theories of truth. This notion can be explicated in many different ways. The most crude is simply via the inclusion of sets of consequences: we may say that  $Th_1$  is not stronger than  $Th_2$ , if  $Th_2$  proves all the axioms of  $Th_1$  ( $Th_2 \vdash Th_1$ ) and  $Th_2$  is strictly stronger, if  $Th_1$  is not stronger than  $Th_2$  but not *vice versa*. Using such a criterion, we easily see, for example, that  $TB$  is not stronger than  $UTB$ : moreover, an easy compactness argument shows that, in fact,  $UTB$  is strictly stronger<sup>11</sup>. Such a definition, in many applications, however, goes against intuition: it classifies some natural theories as incomparable in the situation where one seems to axiomatise a stronger notion of truth than the other. Still, it can be treated as the (obvious, in fact) "upper-bound" on the explications of the notion of strength, in the following sense: if  $Th_2$  proves the axioms of  $Th_1$ , then  $Th_1$  has to be not stronger than  $Th_2$ , according to any reasonable notion of "strength". The next subsection provides us with a more relaxed criterion and the two following subsections are devoted to presenting two notions of strength based on the relation between the truth theory and its base theory. The last one is the main focus of this dissertation, but we think it is worth seeing in a wider context.

#### 3.2.1 Relative Truth Definability

In [17], Kentaro Fujimoto observed that in the context of axiomatic truth theories over the same base theories, the following refinement of the standard relative interpretability relation (as defined in [19], for example) seems to be particularly fruitful:

**Definition 79.** Let  $Th_1$  and  $Th_2$  be two axiomatic truth theories. We say that  $Th_1$  is *relatively truth definable* in  $Th_2$  if there exists a formula  $\phi(x) \in \mathcal{L}_T$  with one free variable, such that for every axiom  $\Psi$  of  $Th_1$

$$Th_2 \vdash \Psi[\phi(x)/T(x)]$$

<sup>11</sup> The intuition behind the argument is that  $TB$  knows nothing about the truth of sentences with non-standard terms. By compactness, it is easy to see that there is a model  $\mathcal{M} \models TB$  such that for some  $c \succ^{\mathcal{M}} \omega$ ,  $\mathcal{M} \models \neg T(\underline{c} = \underline{c})$ . Such a model cannot serve as an interpretation of  $UTB$

where  $\Psi[\phi(x)/T(x)]$  is as explained in Definition 7. We write  $\text{Th}_1 \leq_F \text{Th}_2$ <sup>12</sup> to denote that  $\text{Th}_1$  is relatively truth definable in  $\text{Th}_2$ .

One can base the notion of strength on the above definition in the obvious way:  $\text{Th}_1$  is *not stronger in the sense of Fujimoto* than  $\text{Th}_2$ , if  $\text{Th}_1 \leq_F \text{Th}_2$  and  $\text{Th}_2$  is *stronger in the sense of Fujimoto* than  $\text{Th}_1$  if  $\text{Th}_1$  is not stronger than  $\text{Th}_2$ , but not the other way round (this latter relation will be denoted  $\text{Th}_2 \leq_F \text{Th}_1$ ). The intuition behind this notion of strength is that if  $\text{Th}_1 \leq_F \text{Th}_2$ , then  $\text{Th}_2$  axiomatises a more expressive notion of truth:  $\text{Th}_2$  admits resources to define the notion of truth axiomatised by  $\text{Th}_1$ , but  $\text{Th}_1$  cannot do the same thing with the notion of truth axiomatised by  $\text{Th}_2$ . It is of crucial importance that, in interpreting  $\text{Th}_1$  in  $\text{Th}_2$ , we are allowed neither to redefine notions from the base theory (shared by both truth theories) nor to relativise the range of quantifiers (and such changes are allowed by the standard relation of relative interpretability).

Obviously, if  $\text{Th}_2 \vdash \text{Th}_1$ , then  $\text{Th}_1 \leq_F \text{Th}_2$ , since for the formula  $\phi(x)$  from Definition 79, one can simply take  $T(x)$ , where  $T$  is the truth predicate from  $\text{Th}_2$ . However, there are theories  $\text{Th}_1$  and  $\text{Th}_2$ , such that neither  $\text{Th}_1 \vdash \text{Th}_2$ , nor  $\text{Th}_2 \vdash \text{Th}_1$ , but  $\text{Th}_1 \leq_F \text{Th}_2$ . The examples were already given Fujimoto's original paper: for  $\text{Th}_1$  one can take WKF and for  $\text{Th}_2$ —KF (both theories defined in [17]; the question whether  $\text{WKF} \leq_F \text{KF}$  is still open).

**Remark 80.** It will be worth stretching the scope of Definition 79 to also compare the theories of satisfaction (which can be defined analogously to truth theories, but instead of a unary predicate  $T(x)$  having a binary predicate  $S(x, y)$ ) with other satisfaction or truth theories. The modification, however, is obvious: if  $\text{Th}_1$  and  $\text{Th}_2$  are two theories of satisfaction, then  $\text{Th}_1$  is relatively truth definable in  $\text{Th}_2$  if there exists a formula  $\phi(x, y)$  such that for every axiom  $\Psi$  of  $\text{Th}_1$

$$\text{Th}_2 \vdash \Psi[\phi(x, y)/S(x, y)].$$

The cases when only one of  $\text{Th}_1, \text{Th}_2$  is a theory of satisfaction (and the other a theory of truth) are handled in a similar way.

Loosely speaking, both the above notions of strength are sensitive to the truth theoretical content of the truth theories (in fact, relative truth definability was offered as an explication of precisely this notion) as well as to their *non-semantic content*; i.e. how much do they say about the realm of the base theory. Expressing the last thought a little bit more precisely (in a moment we shall present two possible formal explications of this latter notion), there might be two reasons for  $\text{Th}_1$  not being relatively truth definable in  $\text{Th}_2$ :

1.  $\text{Th}_1$  says more about the notion of truth, but none of the theories delivers more restrictions on the realm of the base theory;
2. In comparison with  $\text{Th}_2$ ,  $\text{Th}_1$  says something essentially new about the realm of the base theory.

---

<sup>12</sup> "F" standing for "Fujimoto".

From the philosophical viewpoint it is worth distinguishing between 1 and 2. For example, if one takes the instrumentalist standpoint towards the notion of truth, then he or she is not interested in what the truth really is, but what consequences accepting a certain theory of truth might have on the *truth-free* part of our knowledge. Once again, the analogy with Reverse Mathematics comes in handy: apart from the questions about which second-order sentences are provable in which systems, we are interested in whether certain axioms have non-finitistic consequences. For example, that WKL is not provable in  $\text{RCA}_0$  is not the end of the story: the interesting thing is that  $\text{RCA}_0$  can be extended with WKL and the resulting theory,  $\text{WKL}_0$  proves the same  $\Pi_1^0$  sentences as  $I\Sigma_1$ <sup>13</sup>. As we already indicated, the motivation standing behind this kind of research in Reverse Mathematics was Hilbert's Programme. A similar role in Axiomatic Truth Theories was played by *Deflationism*, whose central thesis on the "lightness" of truth was proposed to be explicated in terms of *conservativity* over the base theory. From the very beginning, two different notions of conservativity were involved:

**Definition 81** (Conservativity). Let  $\text{Th}$  be an axiomatic theory of truth.

1.  $\text{Th}$  is *proof-theoretically conservative*<sup>14</sup> over PA if for every sentence  $\phi \in \mathcal{L}_{\text{PA}}$ ,

$$\text{Th} \vdash \phi \Rightarrow \text{PA} \vdash \phi.$$

2.  $\text{Th}$  is *model-theoretically conservative*<sup>15</sup> over PA if every model of PA expands (in the sense of Definition 60) to a model of  $\text{Th}$ .

For quite some time, it was claimed<sup>16</sup> that the deflationary theory of truth should be conservative over PA and both notions of conservativity were, in fact, present in the discussion. As for now, we are highly skeptical that this is a correct explication of deflationary theses (mainly after the arguments given in [4]). However, based on both notions of conservativity we can define two notions of strength relative to PA. We start with a stronger one and then proceed to one which is of central interest to this dissertation.

### 3.2.2 Model-theoretical Strength

Let us start with a definition:

**Definition 82.** Let  $\text{Th}_1$  and  $\text{Th}_2$  be two truth theories. We shall say that  $\text{Th}_1$  is *model-theoretically not stronger* than  $\text{Th}_2$ , symbolically  $\text{Th}_1 \leq_M \text{Th}_2$ , if every model of PA which expands to a model of  $\text{Th}_2$ , expands to a model of  $\text{Th}_1$ . We shall say that  $\text{Th}_2$  is *model-theoretically stronger*, symbolically  $\text{Th}_1 \not\leq_M \text{Th}_2$ , if  $\text{Th}_1 \leq_M \text{Th}_2$  but  $\text{Th}_2 \not\leq_M \text{Th}_1$ .

This definition is motivated in the following way: we take the class of models of a theory to be all possibilities admitted by this theory. If there are models of PA that cannot be expanded to

<sup>13</sup> For all these facts and definitions of theories involved, see [41]. The good reference for  $\Pi_1^0$  conservativity of  $\text{WKL}_0$  over  $I\Sigma_1$  is also [47]

<sup>14</sup> This property of theories is sometimes called also "syntactical conservativity".

<sup>15</sup> This property of theories is sometimes called also "semantical conservativity".

<sup>16</sup> See [39], [26], [42] for support of this thesis and [45] for critical remarks.

models of  $\text{Th}$ , it means that  $\text{Th}$  excludes some of the possibilities allowed by  $\text{PA}$ . If  $\text{Th}_2$  is model-theoretically stronger than  $\text{Th}_1$ , then  $\text{Th}_2$  excludes more possibilities than  $\text{Th}_1$ . However, it is to be noted that in most interesting cases the class of possibilities we eliminate cannot be defined by a  $\mathcal{L}_{\text{PA}}$  sentence. Let us give one such example: we shall show that  $\text{TB} \not\leq_M \text{UTB}$ . By Theorem 66 every model of  $\text{UTB}$  is recursively saturated. By Theorem 3.7 in [33] there exists a model of  $\text{TB}$  which is not recursively saturated; hence,  $\text{TB} \not\leq_M \text{UTB}$ . That  $\text{TB} \leq_M \text{UTB}$  is obvious, since  $\text{TB}$  is a subtheory of  $\text{UTB}$ . Since  $\text{TB}$  and  $\text{UTB}$  are both proof-theoretically conservative over  $\text{PA}$ , they do not differ on arithmetical consequences. It is worth mentioning that already  $\text{TB}$  is not model-theoretically conservative over  $\text{PA}$ <sup>17</sup>, hence, the class of  $\text{PA}$  models expandable to  $\text{TB}$  (or  $\text{UTB}$ ) is not closed under  $(\mathcal{L}_{\text{PA}})$  elementary equivalence. In particular, there are models of  $\text{PA}$   $\mathcal{M}_0, \mathcal{M}_1, \mathcal{M}_2$  which are  $(\mathcal{L}_{\text{PA}})$  elementary equivalent, but

1.  $\mathcal{M}_0$  cannot be expanded to a model of  $\text{TB}$ .
2.  $\mathcal{M}_1$  can be expanded to a model of  $\text{TB}$ , but not to a model of  $\text{UTB}$ .
3.  $\mathcal{M}_2$  can be expanded to a model of  $\text{UTB}$ .

### 3.2.3 Proof-Theoretical Strength

Usually, when we talk about proof-theoretical strength of a theory we study how much sentences from certain class  $\Gamma$  the chosen theory proves: for example, many researchers working in Reverse Mathematics concentrate around the question whether a given subsystem of Second Order Arithmetic proves more  $\Pi_1$  sentences than  $I\Sigma_1$  (such consequences are called *non-finitistic*). A theory can be considered "strong" if it has non-finitistic consequences. In Axiomatic Theories of Truth, when investigating the proof-theoretical strength of truth theories, we ask which axioms for the truth predicate allow us to deduce sentences of the base theory that are unprovable by means of the base theory itself. Since the base theory ( $\text{PA}$  in our model) is meant to represent our knowledge on non-semantical facts, the above motivation can be paraphrased in the following way: which properties of the notion of truth allow us to deduce sentences about non-semantical facts that couldn't be deduced from this part of our knowledge which is formulated without invoking the truth predicate? A theory of truth is *strong* if it proves more  $\mathcal{L}_{\text{PA}}$  sentences than  $\text{PA}$  itself. If  $\text{Th}_1$  and  $\text{Th}_2$  are two truth theories, then  $\text{Th}_2$  is stronger if it proves strictly more  $\mathcal{L}_{\text{PA}}$  sentences than  $\text{Th}_1$ . Let us put it in the form of a definition:

**Definition 83.** Let  $\text{Th}_1$  and  $\text{Th}_2$  be two truth theories. We shall say that  $\text{Th}_1$  is *proof-theoretically not stronger*  $\text{Th}_2$ , symbolically  $\text{Th}_1 \leq_P \text{Th}_2$ , if for every sentence  $\phi \in \mathcal{L}_{\text{PA}}$  we have

$$\text{Th}_1 \vdash \phi \Rightarrow \text{Th}_2 \vdash \phi.$$

We shall say that  $\text{Th}_2$  is *proof-theoretically stronger* if  $\text{Th}_1 \leq_P \text{Th}_2$  but  $\text{Th}_2 \not\leq_P \text{Th}_1$ .

In this dissertation, this question will be of particular interest to us: which axiomatic truth theories are strong enough to prove the uniform reflection scheme for  $\text{PA}$  (i.e.  $\mathcal{UR}(\text{PA})$ ), as

<sup>17</sup> This being an observation due to Cezary Cieřliński and Fredrik Engström (independently). For a proof see [33].

defined in Definition 61)? Not all instantiations of this scheme are provable in PA, since for example if

$$\text{Pr}_{\text{PA}}(\ulcorner 0 = 1 \urcorner) \rightarrow 0 = 1$$

was provable, then  $\text{Con}_{\text{PA}}$  would be, contradicting Gödel's Second Incompleteness Theorem (Theorem 77). As the next theorem witnesses, instantiations of reflection with sentences are provable only for those sentences which are already theorems of PA:

**Theorem 84** (Löb's Theorem; [19] Theorem 2.26). *For every formula  $\phi \in \mathcal{L}_{\text{PA}}$  the following are equivalent*

1.  $\text{PA} \vdash \text{Pr}_{\text{PA}}(\ulcorner \phi \urcorner) \rightarrow \phi$ .
2.  $\text{PA} \vdash \phi$ .

In particular, we see that the case of sentence  $0 = 1$  was not exceptional:  $\text{Pr}_{\text{PA}}(\ulcorner \phi \urcorner) \rightarrow \phi$  will be unprovable in PA for every  $\phi$  which is unprovable itself (in particular, for every refutable sentence). However, it is often claimed that PA "implies"  $\mathcal{UR}(\text{PA})$  in a weaker sense. The reflection scheme is claimed to express that consequences of PA are valid, and doing this from the point of view of PA, since  $\text{Pr}_{\text{PA}}$  predicate naturally formalises provability in PA. Hence, if anyone accepts PA and admits that inferences in First-Order Logic preserve validity, then he or she should accept  $\mathcal{UR}(\text{PA})$ . Such a view has been discussed, for example, in [22]. Feferman's classical paper ([11]) studies which theories are obtained as limit of this process (see also [8] for an overview and [16] for a gentle introduction). We have to admit that there is a lot of philosophical work to clarify these intuitions (see [4]), but the intuitions seems correct. We think that the results of this dissertation shed some light on which truth principles allow us to deduce the implicit commitments of PA.

### 3.2.4 Relations between the three notions of strength

Before we introduce the truth theories we would like to study, let us make explicit the relations between the three above notions of strength.

**Proposition 85.** *For every axiomatic truth theories  $\text{Th}_1$  and  $\text{Th}_2$ :*

$$\text{Th}_1 \leq_F \text{Th}_2 \Rightarrow \text{Th}_1 \leq_M \text{Th}_2 \Rightarrow \text{Th}_1 \leq_P \text{Th}_2.$$

*Sketch of the proof.* We prove  $\text{Th}_1 \leq_F \text{Th}_2 \Rightarrow \text{Th}_1 \leq_M \text{Th}_2$  first. Assume  $\text{Th}_1 \leq_F \text{Th}_2$  and let  $\phi(x)$  define  $\text{Th}_1$  in  $\text{Th}_2$ . Let us pick arbitrary model  $\mathcal{M} \models \text{PA}$  which expands to  $\text{Th}_2$  and let  $\mathcal{M}' := (\mathcal{M}, Tr)$  be the result of this expansion. Then  $(\mathcal{M}, \phi^{\mathcal{M}'})$  is a model of  $\text{Th}_1$  (we use Convention 8). To prove  $\text{Th}_1 \leq_M \text{Th}_2 \Rightarrow \text{Th}_1 \leq_P \text{Th}_2$  assume the former holds. Pick arbitrary  $\phi \in \mathcal{L}_{\text{PA}}$  such that  $\text{Th}_1 \vdash \phi$ . To see that  $\text{Th}_2 \vdash \phi$  fix arbitrary model  $\mathcal{M} \models \text{PA}$  such that  $(\mathcal{M}, Tr) \models \text{Th}_2$ . Since  $\mathcal{M}$  expands to a model of  $\text{Th}_2$ , then it expands to a model of  $\text{Th}_1$ , hence,  $\mathcal{M} \models \phi$ .  $\square$

It is worth mentioning that none of the above implications reverses: we have already said that TB and UTB have the same consequences in  $\mathcal{L}_{PA}$ , but  $UTB \not\leq_M TB$ . Their non-inductive counterparts provide a counter example to the implication  $Th_1 \leq_M Th_2 \Rightarrow Th_1 \leq_F Th_2$ :  $UTB^-$  and  $TB^-$  are model-theoretically conservative over PA<sup>18</sup>, but  $UTB^- \not\leq_F TB^-$ . The latter holds, for otherwise we would have  $UTB \leq_F TB$ <sup>19</sup> and this is not true by the first part of the proof and already mentioned fact that  $UTB \not\leq_M TB$ . Let us state one more proposition, that can be proved in a similar way to that stated above:

**Proposition 86** ([17]). *Let  $Th_1$  and  $Th_2$  be two theories of truth (or satisfaction) such that  $Th_1 \leq_F Th_2$ . Then*

1. *if  $Th_2$  is proof-theoretically (model-theoretically) conservative over PA, then  $Th_1$  is proof-theoretically (model-theoretically) conservative over PA.*
2. *if  $Th_1$  is not proof-theoretically (model-theoretically) conservative over PA, then  $Th_2$  is not proof-theoretically (model-theoretically) conservative over PA.*

This dissertation is oriented mainly on proof-theoretical strength. However, at some points we shall mention the other two notions. Lastly, let us observe that having three different notions of *strength* of axiomatic theories of truth makes it possible to measure not only whether one theory is stronger than another one, but also how much stronger it is. In the next section we define the theories which will be of interest.

### 3.3 Definitions of Axiomatic Theories of Truth in Study

In this section we introduce all the basic theories studied in the current dissertation. All the theories will be *stratified*, i.e. the notion of truth is characterized only for the arithmetised language  $\mathcal{L}_{PA}$ . Although the main focus of this thesis is axiomatic theories of *truth*, in some contexts (for example, in model theory of Peano Arithmetic) the notion of *satisfaction* plays a more important role. Moreover, this notion can be generalised to theories which does not prove that every object can be named by a closed term (ZFC being the most important example). Last but not least axiomatic theories of satisfaction are, in general, relatively truth definable (made precise in Subsection 3.2.1, Remark 80) in their truth theoretic counterparts and in this sense, are weaker. To sum up: axiomatic theories of satisfaction enable us to obtain more general results. Below, we introduce all the relevant theories of truth and their "satisfaction counterparts". We start with some handy notational conventions:

**Definition 87** (PA).

1. If  $x$  is a term or a formula, then

<sup>18</sup> The proof is very easy and moreover, both theories are subtheories of  $PT^-$  and  $WPT^-$ , hence, it is also a consequence of results mentioned in Subsection 3.3.4 and Theorem 209

<sup>19</sup> As can be easily verified, if  $Th_1$  is relatively truth definable in  $Th_2$ , then  $Th_1 + \text{Ind}(\mathcal{L}_T)$  is relatively truth definable in  $Th_2 + \text{Ind}(\mathcal{L}_T)$ . See also Proposition 28 in [17].

- (a) we say that an assignment  $\alpha$  is *an assignment for  $x$*  if and only if  $\text{dom}(\alpha) = \text{FV}(x)$  ( $\text{dom}(\alpha) = \text{Var}(x)$ , if  $x$  is a term). The set of assignments for  $x$  will be denoted by  $\text{Asn}(x)$ .
- (b) If  $c$  is any set of terms or variables, then  $\text{Asn}(c) = \bigcup_{x \in c} \text{Asn}(x)$ . If  $x_0, \dots, x_n$  are any terms or formulae, then  $\text{Asn}(x_0, \dots, x_n)$  is a short for  $\text{Asn}(\{x_0, \dots, x_n\})$ .
2. If  $\alpha, \beta$  are assignments and  $v$  is a variable then  $\beta \sim_v \alpha$  means that  $\beta$  differs from  $\alpha$  at most on the value assigned to the variable  $v$  and  $\beta$  is *defined* on  $v$  variable. This includes the situation in which  $\alpha$  does not assign anything to  $v$  and  $\beta$  assigns something to it. We note that, in contrast with the logical tradition, this relation is not an equivalence relation. We prefer such format because it will be more convenient later on.
3. if  $\phi$  is a formula and  $\alpha$  is an assignment, then  $\alpha \upharpoonright_\phi$  denotes the restriction of assignment  $\alpha$  to the free variables of  $\phi$ . In the context of a satisfaction predicate  $S$ ,  $S(\phi, \alpha \upharpoonright_\cdot)$  abbreviates  $S(\phi, \upharpoonright_\phi)$ .

### 3.3.1 Classically Compositional Theories

**Definition 88** ( $\text{CT}^-$ ).  $\text{CT}^-$  is the  $\mathcal{L}_T$  theory extending PA with the following axioms:

1.  $\forall x (T(x) \rightarrow \text{Sent}_{\mathcal{L}_{\text{PA}}}(x))$
2.  $\forall s, t (T(s = t) \equiv (s^\circ = t^\circ))$
3.  $\forall \phi (T(\neg\phi) \equiv \neg T(\phi))$
4.  $\forall \phi, \psi (T(\phi \vee \psi) \equiv (T(\phi) \vee T(\psi)))$
5.  $\forall v \forall \phi(v) (T(\exists v \phi) \equiv \exists x T(\phi(\underline{x})))$

**Proposition 89** (Truth conditions for  $\wedge$ ,  $\rightarrow$  and  $\forall$ ). *The following sentences are provable in  $\text{CT}^-$*

1.  $\forall \phi, \psi (T(\phi \wedge \psi) \equiv (T(\phi) \wedge T(\psi)))$
2.  $\forall \phi, \psi (T(\phi \rightarrow \psi) \equiv (T(\phi) \rightarrow T(\psi)))$
3.  $\forall v \forall \phi(v) (T(\forall v \phi) \equiv \forall x T(\phi(\underline{x})))$

**Definition 90** ( $\text{CS}^-$ ).  $\text{CS}^-$  is the extension of PA with the following axioms:

1.  $\forall x, y (S(x, y) \rightarrow x \in \text{Form}_{\mathcal{L}_{\text{PA}}} \wedge y \in \text{Asn}(x))$ .
2.  $\forall s(\bar{x}), t(\bar{x}) \forall \alpha \in \text{Asn}(s, t) (S(s = t, \alpha) \equiv (s)_\alpha^\circ = (t)_\alpha^\circ)$ .
3.  $\forall \phi(\bar{x}), \psi \forall \alpha \in \text{Asn}(\phi, \psi) (S(\phi \vee \psi, \alpha) \equiv S(\phi, \alpha \upharpoonright_\cdot) \vee S(\psi, \alpha \upharpoonright_\cdot))$ .

4.  $\forall\phi(\bar{x})\forall\alpha \in \text{Asn}(\phi) \left( S(\neg\phi, \alpha) \equiv \neg S(\phi, \alpha) \right)$ .
5.  $\forall\phi(\bar{x})\forall v\forall\alpha \in \text{Asn}(\exists v\phi) \left( S(\exists v\phi, \alpha) \equiv \exists\beta \sim_v \alpha \ S(\phi, \beta \uparrow) \right)$ .

Recall that in the arithmetised language  $\mathcal{L}_{\text{PA}}$ , for all formulae  $\phi, \psi, \phi \wedge \psi, \phi \rightarrow \psi$  and  $\forall v\phi$  are treated as abbreviations of  $\neg(\neg\phi \vee \neg\psi)$ ,  $\neg\phi \vee \psi$  and  $\neg\exists v\neg\phi$ , respectively. Also we have the following easily provable proposition

**Proposition 91** (Satisfaction conditions for  $\wedge, \rightarrow$  and  $\forall$ ). *The following sentences are provable in  $\text{CS}^-$*

1.  $\forall\phi(\bar{x}), \psi(\bar{x}), \forall\alpha \in \text{Asn}(\phi, \psi) \left( S(\phi \wedge \psi, \alpha) \equiv S(\phi, \alpha \uparrow_\phi) \wedge S(\psi, \alpha \uparrow_\psi) \right)$ .
2.  $\forall\phi(\bar{x}), \psi(\bar{x}), \forall\alpha \in \text{Asn}(\phi, \psi) \left( S(\phi \rightarrow \psi, \alpha) \equiv S(\phi, \alpha \uparrow_\phi) \rightarrow S(\psi, \alpha \uparrow_\psi) \right)$ .
3.  $\forall\phi(\bar{x})\forall v\forall\alpha \in \text{Asn}(\forall v\phi) \left( S(\forall v\phi, \alpha) \equiv \forall\beta \sim_v \alpha \ S(\phi, \beta \uparrow_\phi) \right)$ .

**Remark 92** (Full satisfaction classes in models of PA). If  $\mathcal{M}$  is a model of PA then any  $A \subset M^2$  such that

$$(\mathcal{M}, A) \models \text{CS}^-$$

is called a *full satisfaction class* on  $\mathcal{M}$ . This notion is very important in the study of structure of models of PA. Its significance is described, for example, in [24] and [27].

**Definition 93** (CT and its subsystems). CT is the theory axiomatised by  $\text{CT}^-$  and  $\text{Ind}(\mathcal{L}_T)$  and  $\text{CT}_n$  is the theory axiomatised by  $\text{CT}^-$  and  $I\Sigma_n(\mathcal{L}_T)$ .

### 3.3.2 Non-Classically Compositional Theories of Truth

To our best knowledge theory  $\text{PT}^-$  has been introduced in [12] as the stratified counterpart of Kripke-Feferman theory  $\text{KF}^-$  (defined as  $\text{KF} \uparrow$  in [20]). Its introduction was connected with the search of axiomatic theory of truth that would be both semantically conservative over PA and not relatively interpretable in it.

**Definition 94** ( $\text{PT}^-$ ).  $\text{PT}^-$  is the  $\mathcal{L}_T$  theory extending PA with the following axioms:

1.  $\forall x \left( T(x) \rightarrow \text{Sent}_{\mathcal{L}_{\text{PA}}}(x) \right)$
2.  $\forall\phi \left( T(\neg\neg\phi) \equiv T(\phi) \right)$
3. (a)  $\forall s, t \left( T(s = t) \equiv (s^\circ = t^\circ) \right)$   
(b)  $\forall s, t \left( T(\neg(s = t)) \equiv (s^\circ \neq t^\circ) \right)$
4. (a)  $\forall\phi, \psi \left( T(\phi \vee \psi) \equiv (T(\phi) \vee T(\psi)) \right)$   
(b)  $\forall\phi, \psi \left( T(\neg(\phi \vee \psi)) \equiv (T(\neg\phi) \wedge T(\neg\psi)) \right)$

5. (a)  $\forall v \forall \phi(v) \left( T(\exists v \phi) \equiv \exists x T(\phi(\underline{x})) \right)$   
 (b)  $\forall v \forall \phi(v) \left( T(\neg \exists v \phi) \equiv \forall x T(\neg \phi(\underline{x})) \right)$

In  $PT^-$  (and  $KF^-$ ), the internal logic is modelled after Strong Kleene Logic. In [17] theory  $WKF^-$  has been introduced: the internal logic of this theory is modelled after Weak Kleene Logic. Theory  $WPT^-$  introduced below is its natural stratified counterpart. Before giving its axioms, we shall introduce formulae that are crucial in studying properties of non-classically compositional theories:

**Definition 95.**

$$\begin{aligned} \text{tot}(\phi) &:= \text{Form}_{\mathcal{L}_{PA}}^{\leq 1}(\phi) \wedge \forall x (T(\phi(\underline{x})) \vee T(\neg \phi(\underline{x}))) \\ \text{cons}(\phi) &:= \text{Form}_{\mathcal{L}_{PA}}^{\leq 1}(\phi) \wedge \neg \exists x (T(\phi(\underline{x})) \wedge T(\neg \phi(\underline{x}))) \\ \text{Tot} &:= \forall v \forall \phi(v) \text{tot}(\phi) \\ \text{Cons} &:= \forall v \forall \phi(v) \text{cons}(\phi) \end{aligned}$$

Moreover to save place, let us introduce the following abbreviation (which gives the semantics for the disjunction according to the Weak Kleene Logic):

$$\vee_{wk}(\phi, \psi) := (T(\phi) \wedge T(\psi)) \vee (T(\neg \phi) \wedge T(\psi)) \vee (T(\phi) \wedge T(\neg \psi))$$

**Definition 96** ( $WPT^-$ ).  $WPT^-$  is the  $\mathcal{L}_T$  theory extending PA with axioms 1, 2, 3(a), 3(b), 4(b), 5(b) from Definition 94 and the following sentences:

- 4(a)<sub>w</sub>  $\forall \phi, \psi \left( T(\phi \vee \psi) \equiv \vee_{wk}(\phi, \psi) \right)$   
 5(a)<sub>w</sub>  $\forall v \forall \phi(v) \left( T(\exists v \phi) \equiv (\text{tot}(\phi) \wedge \exists x T(\phi(\underline{x}))) \right)$

**Proposition 97.** *The following sentences are provable in both  $PT^-$  and  $WPT^-$ :*

1.  $\forall \phi, \psi \left( T(\phi \wedge \psi) \equiv (T(\phi) \wedge T(\psi)) \right)$
2.  $\forall v \forall \phi(v) \left( T(\forall v \phi) \equiv \forall x T(\phi(\underline{x})) \right)$
3.  $\forall \phi, \psi \left( \vee_{wk}(\phi, \psi) \equiv (\text{tot}(\phi) \wedge \text{tot}(\psi) \wedge T(\phi) \vee T(\psi)) \right)$

Moreover, the following sentences are provable in  $PT^-$ :

4.  $\forall \phi, \psi \left( T(\neg(\phi \wedge \psi)) \equiv (T(\neg \phi) \vee T(\neg \psi)) \right)$
5.  $\forall \phi, \psi \left( \text{tot}(\phi) \wedge \text{cons}(\phi) \longrightarrow T(\phi \rightarrow \psi) \equiv (T(\phi) \rightarrow T(\psi)) \right)$
6.  $\forall v \forall \phi(v) \left( T(\neg \forall v \phi) \equiv \exists x T(\neg \phi(\underline{x})) \right)$

and the following are provable in  $\text{WPT}^-$ :

7.  $\forall \phi, \psi \left( T(\neg(\phi \wedge \psi)) \equiv (\text{tot}(\phi) \wedge \text{tot}(\psi) \wedge (T(\neg\phi) \vee T(\neg(\psi)))) \right)$
8.  $\forall \phi, \psi \left( \text{tot}(\phi) \wedge \text{cons}(\phi) \wedge \text{tot}(\psi) \longrightarrow T(\phi \rightarrow \psi) \equiv (T(\phi) \rightarrow T(\psi)) \right)$
9.  $\forall v \forall \phi(v) \left( T(\neg \forall v \phi) \equiv \text{tot}(\phi) \wedge \exists x T(\phi(x)) \right)$

**Proposition 98.**  $\text{CT}^-$  proves the axioms of both  $\text{PT}^-$  and  $\text{WPT}^-$ . Moreover

1.  $\text{PT}^- + \text{Tot} \vdash \text{WPT}^-$
2.  $\text{WPT}^- + \text{Tot} \vdash \text{PT}^-$
3.  $\text{PT}^- + \text{Tot} + \text{Cons} \vdash \text{CT}^-$
4.  $\text{WPT}^- + \text{Tot} + \text{Cons} \vdash \text{CT}^-$

**Proposition 99.** For every formula of  $\mathcal{L}_{\text{PA}}$   $\phi(x_0, \dots, x_n)$  both  $\text{PT}^-$  and  $\text{WPT}^-$  prove

$$\forall t_0, \dots, t_n \left( T(\phi(t_0, \dots, t_n)) \equiv \phi(t_0^\circ, \dots, t_n^\circ) \right)$$

In particular, every standard formula of  $\mathcal{L}_{\text{PA}}$  is total and consistent, provably in both theories.

Let us now define two non-classically compositional theories of satisfaction. To our best knowledge this is the first time these theories have become objects of study.

**Definition 100** ( $\text{PS}^-$ ).  $\text{PS}^-$  is the extension of PA with the following axioms:

1.  $\forall x, y (S(x, y) \rightarrow x \in \text{Form}\mathcal{L}_{\text{PA}} \wedge y \in \text{Asn}(x))$ .
2.  $\forall \phi(\bar{x}) \forall \alpha \in \text{Asn}(\phi) \left( S(\neg\neg\phi, \alpha) \equiv S(\phi, \alpha) \right)$ .
3. (a)  $\forall s, t \in \text{Term} \forall \alpha \in \text{Asn}(s, t) (S(s = t, \alpha) \equiv (s)_\alpha^\circ = (t)_\alpha^\circ)$ .  
(b)  $\forall s, t \in \text{Term} \forall \alpha \in \text{Asn}(s, t) (S(\neg s = t, \alpha) \equiv (s)_\alpha^\circ \neq (t)_\alpha^\circ)$ .
4. (a)  $\forall \phi(\bar{x}), \psi(\bar{x}) \forall \alpha \in \text{Asn}(\phi, \psi) \left( S(\phi \vee \psi, \alpha) \equiv S(\phi, \alpha \upharpoonright) \vee S(\psi, \alpha \upharpoonright) \right)$ .  
(b)  $\forall \phi(\bar{x}), \psi(\bar{x}) \forall \alpha \in \text{Asn}(\phi, \psi) \left( S(\neg(\phi \vee \psi), \alpha) \equiv S(\neg\phi, \alpha \upharpoonright) \wedge S(\neg\psi, \alpha \upharpoonright) \right)$ .
5. (a)  $\forall \phi(\bar{x}) \forall v \forall \alpha \in \text{Asn}(\exists v \phi) \left( S(\exists v \phi, \alpha) \equiv \exists \beta \sim_v \alpha S(\phi, \beta \upharpoonright) \right)$ .  
(b)  $\forall \phi(\bar{x}) \forall v \forall \alpha \in \text{Asn}(\exists v \phi) \left( S(\neg \exists v \phi, \alpha) \equiv \forall \beta \sim_v \alpha S(\neg\phi, \beta \upharpoonright) \right)$ .

The following definition generalises the notion of total formulae in theories of satisfaction:

**Definition 101.**

$$\begin{aligned} \text{tot}(\phi, \alpha) &:= S(\phi, \alpha) \vee S(\neg\phi, \alpha) \\ \text{tot}_v(\phi, \alpha) &:= \forall \beta \sim_v \alpha \text{tot}(\phi, \beta \upharpoonright). \\ \text{Tot}(S) &:= \forall \phi(\bar{x}) \forall \alpha \in \text{Asn}(\phi) \text{tot}(\phi, \alpha). \end{aligned}$$

For further usage let us define also the dual of  $\text{tot}_v(\alpha, \phi)$ :

**Definition 102.**

$$\begin{aligned}\text{cons}(\phi, \alpha) &:= \neg(S(\phi, \alpha) \wedge S(\neg\phi, \alpha)) \\ \text{cons}_v(\phi, \alpha) &:= \forall\beta \sim_v \alpha \text{ cons}(\phi, \beta). \\ \text{Cons}(S) &:= \forall\phi(\bar{x})\forall\alpha \in \text{Asn}(\phi) \text{ cons}(\phi, \alpha).\end{aligned}$$

Moreover, to save place, let us use the following abbreviation (analogous to  $\vee_{wk}(\phi, \psi)$  defined earlier):

$$\begin{aligned}\vee_{wk}(\phi, \psi, \alpha) &:= \left( (S(\phi, \alpha \upharpoonright_\phi) \wedge S(\psi, \alpha \upharpoonright_\psi)) \vee (S(\neg\phi, \alpha \upharpoonright_\phi) \wedge S(\psi, \alpha \upharpoonright_\psi)) \right. \\ &\quad \left. \vee (S(\phi, \alpha \upharpoonright_\phi) \wedge S(\neg\psi, \alpha \upharpoonright_\psi)) \right)\end{aligned}$$

$\vee_{wk}$  gives the semantics for disjunction according to Weak Kleene Logic.

**Definition 103 (WPS<sup>-</sup>).** WPS<sup>-</sup> is the extension of PA with axioms 1, 2, 3, 4(b), 5(b) from the definition of PS<sup>-</sup> and additionally

$$4(a)_w \quad \forall\phi(\bar{x}), \psi(\bar{x})\forall\alpha \in \text{Asn}(\phi, \psi) \quad (S(\phi \vee \psi, \alpha) \equiv \vee_{wk}(\phi, \psi, \alpha)).$$

$$5(a)_w \quad \forall\phi(\bar{x})\forall v\forall\alpha \in \text{Asn}(\exists v\phi) \quad \left( S(\exists v\phi, \alpha) \equiv \text{tot}_v(\phi, \alpha) \wedge \exists\beta \sim_v \alpha \ S(\phi, \beta \upharpoonright_\phi) \right).$$

As usual  $\phi \wedge \psi$  in the arithmetised language abbreviates  $\neg(\neg\phi \vee \neg\psi)$  and  $\forall v\phi$  abbreviates  $\neg\exists v\neg\phi$ . For completeness, let us note that both theories prove the expected truth conditions for  $\wedge$  and  $\forall$ :

**Proposition 104.** *If Th is any of PS<sup>-</sup>, WPS<sup>-</sup>, then the following hold*

1.  $\text{Th} \vdash \forall\phi(\bar{x}), \psi(\bar{x})\forall\alpha \in \text{Asn}(\phi, \psi) \quad (S(\phi \wedge \psi, \alpha) \equiv S(\phi, \alpha \upharpoonright_\phi) \wedge S(\psi, \alpha \upharpoonright_\psi))$
2.  $\text{Th} \vdash \forall\phi(\bar{x})\forall v\forall\alpha \in \text{Asn}(\forall v\phi) \quad (S(\forall v\phi, \alpha) \equiv \forall\beta \sim_v \alpha \ S(\phi, \beta \upharpoonright_\phi))$

Moreover, the following sentences are provable in WPS<sup>-</sup>:

3.  $\forall\phi(\bar{x}), \psi(\bar{x})\forall\alpha \in \text{Asn}(\phi, \psi) \quad (S(\neg(\phi \wedge \psi), \alpha) \equiv (\vee_{wk}(\neg\phi, \neg\psi, \alpha)))$
4.  $\forall\phi(\bar{x})\forall v\forall\alpha \in \text{Asn}(\forall v\phi) \quad (S(\neg\forall v\phi, \alpha) \equiv (\text{tot}_v(\phi, \alpha) \wedge \exists\beta \sim_v \alpha \ S(\neg\phi, \beta \upharpoonright_\phi)))$

and the following are provable in PS<sup>-</sup>:

5.  $\forall\phi(\bar{x}), \psi(\bar{x})\forall\alpha \in \text{Asn}(\phi, \psi) \quad (S(\neg(\phi \wedge \psi), \alpha) \equiv (S(\neg\phi, \alpha \upharpoonright_\phi) \vee S(\neg\psi, \alpha \upharpoonright_\psi)))$
6.  $\forall\phi(\bar{x})\forall v\forall\alpha \in \text{Asn}(\forall v\phi) \quad (S(\neg\forall v\phi, \alpha) \equiv (\exists\beta \sim_v \alpha \ S(\neg\phi, \beta \upharpoonright_\phi)))$

Let us make the following easy observations, analogous to those made while introducing CT<sup>-</sup>, PT<sup>-</sup> and WPT<sup>-</sup>:

**Proposition 105.**  $CS^-$  proves the axioms of both  $PS^-$  and  $WPS^-$ . Moreover

1.  $PS^- + \text{Tot}(S) \vdash WPS^-$
2.  $WPS^- + \text{Tot}(S) \vdash PS^-$
3.  $PS^- + \text{Tot}(S) + \text{Cons}(S) \vdash CS^-$
4.  $WPS^- + \text{Tot}(S) + \text{Cons}(S) \vdash CS^-$

The following definition is analogous that given for  $CT^-$ :

**Definition 106.**  $PT$  and  $WPT$  denote the fully inductive extensions of  $PT^-$  and  $WPT^-$ , respectively. For every  $n$   $PT_n$  and  $WPT_n$  denote the extensions of  $PT^-$  and  $WPT^-$  with instantiations of induction scheme for  $\Sigma_n \mathcal{L}_T$  formulae.  $PS, PS_n, WPS, WPS_n$  are analogously defined extensions of  $PS^-$  and  $WPS^-$ .

The next proposition gives a method of relatively truth defining theories of satisfaction in respective theories of truth. We state it for the pair  $CT^-$  and  $CS^-$  but it holds without changes, also for the pairs

$$\begin{aligned} &PT^- \text{ and } PS^- \\ &WPT^- \text{ and } WPS^- \end{aligned}$$

**Proposition 107.** Let  $\phi_S(x, y)$  be the  $\mathcal{L}_T$  formula

$$\text{Form}_{\mathcal{L}_{PA}}(x) \wedge y \in \text{Asn}(x) \wedge T(x[y])$$

Then  $CT^- \vdash CS^-[\phi_S(x, y)/S(x, y)]$  and  $CT_0 \vdash CS_0[\phi_S(x, y)/S(x, y)]$

*Proof.* The only non-obvious part is to show that  $CT_0 \vdash \Delta_0(\phi_S(x, y))$ . By Proposition 56 it is enough to demonstrate that in each model  $\mathcal{M}$  of  $CT_0$ ,  $\phi_S^{\mathcal{M}}(x, y)$  is a class (in the sense of Definition 55). So let us fix arbitrary  $\mathcal{M} \models CT_0$  and  $c \in M$ . By the collection principle in  $\mathcal{M}$  there is  $d$  such that for every  $x, y$  such that  $\langle x, y \rangle < c$ , if  $x$  is a formula and  $y$  is an assignment for  $y$ ,  $x[y]$  is less than  $d$ . Let  $e$  be such that

$$\mathcal{M} \models \forall z < d \ (T(z) \equiv z \in e)$$

(the existence of such an  $e$  follows from the fact that  $\mathcal{M} \models \Delta_0(T)$ ). Then in  $\mathcal{M}$  it holds that for all  $x, y$  such that  $\langle x, y \rangle < c$

$$\phi_S(x, y) \equiv (\text{Form}_{\mathcal{L}_{PA}}(x) \wedge y \in \text{Asn}(x) \wedge x[y] \in e)$$

Since the formula on the right-hand side is arithmetical we may find the element coding the set

$$\{\langle x, y \rangle \in M \mid \langle x, y \rangle < c \wedge \mathcal{M} \models \text{Form}_{\mathcal{L}_{PA}}(x) \wedge y \in \text{Asn}(x) \wedge x[y] \in e\}$$

which suffices by the above equivalence. □

It might seem that the converse to the above theorem should be obvious, as well: working in  $\text{Th}$ , which can be any of  $\text{CS}^-$ ,  $\text{PS}^-$ ,  $\text{WPS}^-$ , the formula

$$T(x) := S(x, \varepsilon)$$

should serve as an interpretation of the truth predicate of the respective theory of truth. This, however, is not true for theories without induction and problematic for  $\Delta_0$  inductive ones. The delicate issue is that, having so restricted means, it is impossible (in the former case) or not obvious (in the latter case) that for all  $\phi(v)$  and all  $x$  we have

$$S(\phi(v), [v \mapsto x]) \equiv S(\phi(\underline{x}), \varepsilon)$$

which is needed to prove the quantifier axiom of the respective theory of truth ( $[v \mapsto x]$  denotes the assignment sending  $v$  to  $x$  and undefined for the rest of variables). The fact that the above is not provable in  $\text{CS}^-$  can be demonstrated using Enayat-Visser techniques from [9]. In Chapter 4, we shall show that the above is provable in  $\text{CS}_0$  and later on, in Section 5.1, that it holds also in  $\text{PS}_0$  and a strengthening of  $\text{WPS}_0$ .

### 3.3.3 Strength of Classically Compositional Theories

Let us now present what is already known concerning the strength of classically compositional theories. The most important theorem states that such theories without induction are proof-theoretically *weak*. By Proposition 107 and Proposition 86, it is enough to state it for  $\text{CT}^-$ .

**Theorem 108** (Krajewski-Kotlarski-Lachlan[29], Enayat-Visser[9], Leigh[31]).  *$\text{CT}^-$  is proof-theoretically conservative over PA.*

However, both theories,  $\text{CT}^-$  and  $\text{CS}^-$  are not completely innocent: from the class of all models of PA they eliminate all non-standard models that are not recursively saturated. The best source to the theorem below, witnessing semantical strength of  $\text{CS}^-$  is [24], Theorem 15.5. Once again, the analogous theorem for  $\text{CT}^-$  is an easy consequence, since  $\text{CS}^-$  is relatively interpretable in  $\text{CT}^-$ .

**Theorem 109** (Lachlan). *Every model of  $\text{CS}^-$  is recursively saturated. In particular,  $\text{CS}^-$  is not model-theoretically conservative over PA.*

Putting together the above theorems and what was said about TB and UTB, we see three very different theories of truth which share arithmetical consequences with PA (are proof-theoretically weak), but all eliminate some models (are model-theoretically strong). It turned out that they can be compared with respect to model-theoretical strength. The theorem below can be seen as strengthening of Lachlan's theorem, since it is a routine argument that every model of UTB is recursively saturated.

**Theorem 110** (Weisło, [33]).  *$\text{UTB} \leq_M \text{CT}^-$ .*

Let us now pass to the extensions of  $\text{CT}^-$  with induction scheme for formulae with the truth predicate. Since we know that every such extension is not model-theoretically conservative over PA, in the rest of this subsection we shall often use "conservative" to mean "proof-theoretically conservative over PA". We have the following folklore result:

**Theorem 111** ([20]). *CT is not proof-theoretically conservative over PA.*

A natural question arises: which extensions of  $CT^-$  are conservative? The line demarcating its conservative extensions from the non-conservative ones has been called *the Tarski Boundary*<sup>20</sup>. In fact, the usual proof of Theorem 111 proceeds by demonstrating that the following principle, called *Global Reflection*, is provable in CT:

$$\forall\phi \ (\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi)). \quad (\text{REF})$$

Obviously each extension of  $CT^-$ , which proves (REF) is not conservative. Since CT is not finitely axiomatisable<sup>21</sup> the above sentence has to be provable in one of  $CT_n$ 's. An inspection of the proof quickly shows that

$$CT_1 \vdash \forall\phi \ (\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi)).$$

In principle, the Tarski Boundary must pass below  $CT_1$ . In [28] Henryk Kotlarski presented a proof of the following theorem:

**Theorem 112.**  $CS_0 \vdash \forall\phi(\text{Pr}_{\text{PA}}(\phi) \rightarrow \forall\beta \in \text{Asn}(\phi)S(\phi, \beta))$

The above theorem would imply that also  $CT_0$  lies on the non-conservative side of the Tarski Boundary. The proof was very sketchy, but suggestive and similar arguments (with references to Kotlarski) were given also in [7] and [23]. The idea was very straightforward indeed: fix  $\phi$  and let  $\phi_0, \dots, \phi_a$  be any proof of  $\phi$  in PA. Let  $\beta \in \text{Asn}(\phi_0, \dots, \phi_a)$ . Use a version of proof system with modus ponens as the only rule of reasoning (as the one we actually defined in the last section). By induction on the length of proof, show that for every  $x \leq a$   $S(\phi_x, \beta \upharpoonright_{\phi_x})$ . The inductive assumption is a  $\Delta_0$  formula because the size of each  $\phi_x$  can be bounded by the code of our proof. What is left to check is that all axioms of PA and First Order Logic are true. This is where the problems begin: it is not clear at all that the notion of satisfaction as axiomatised by  $CS_0$  is sufficiently well-behaved to prove this. Albert Visser and Richard Heck first noticed the problem (independently) around 2008 and since then, even the question of conservativity of  $CS_0$  over PA was considered as an open question. This is because  $\Delta_0$  induction does not allow to freely use induction on the build-up of formulae, since the step for quantifiers requires (at first glance)  $\Pi_1$  induction. Bartosz Wcisło first showed that  $CT_0$  is a strong theory. More precisely, he demonstrated the following theorem:

**Theorem 113** (Wcisło, [46]).  $CT_0 + \forall\phi \ (\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi))$  is relatively truth definable in  $CT_0$ .

Moreover, his proof can be adapted to the case of  $CS^-$ . This still left open the problem of whether Kotlarski's claim is true and whether  $CT_0$  is strong enough to actually prove (REF). In section 4 we show how to fix Kotlarski's original proof.

<sup>20</sup> Name proposed by Ali Enayat.

<sup>21</sup> Being fully inductive, it cannot be for the same reasons as PA, see [24], [19].

## 3.3.4 Strength of Non-Classically Compositional Theories

Let us start by stressing that both  $\text{PT}^-$  and  $\text{WPT}^-$  are model-theoretically weaker than  $\text{CT}^-$ :

**Theorem 114.**  $\text{PT}^-$  and  $\text{WPT}^-$  are model-theoretically conservative over PA.

We shall sketch the standard argument for  $\text{PT}^-$ . Various its modifications for stronger theories will be extensively used in Section 5. In particular, Theorem 209 will imply that  $\text{WPT}^-$  is model-theoretically conservative over PA. Let us recall the standard definition of a function which generates possible extensions for the  $\text{PT}^-$  truth predicate.

**Definition 115.** Let  $\mathcal{M} \models \text{PA}$ .

$$\begin{aligned} \Theta_{\mathcal{M}}(\phi, A) := & \mathcal{M} \models \exists s, t [\phi = (s = t) \wedge s^\circ = t^\circ] \\ & \vee \mathcal{M} \models \exists s, t [\phi = \neg(s = t) \wedge s^\circ \neq t^\circ] \\ & \vee \exists \psi \in \text{Sent}_{\mathcal{L}_{\text{PA}}}^{\mathcal{M}} [\mathcal{M} \models \phi = \neg\neg\psi] \wedge \psi \in A \\ & \vee \exists \psi_1, \psi_2 \in \text{Sent}_{\mathcal{L}_{\text{PA}}}^{\mathcal{M}} [\mathcal{M} \models \phi = (\psi_1 \vee \psi_2)] \wedge (\psi_1 \in A) \vee (\psi_2 \in A) \\ & \vee \exists \psi_1, \psi_2 \in \text{Sent}_{\mathcal{L}_{\text{PA}}}^{\mathcal{M}} [\mathcal{M} \models \phi = \neg(\psi_1 \vee \psi_2)] \wedge (\neg\psi_1 \in A) \wedge (\neg\psi_2 \in A) \\ & \vee \exists \psi \in \text{Form}_{\mathcal{L}_{\text{PA}}}^{\mathcal{M}} [\mathcal{M} \models \phi = \exists x\psi] \wedge \exists s (\psi(\underline{s}) \in A) \\ & \vee \exists \psi \in \text{Form}_{\mathcal{L}_{\text{PA}}}^{\mathcal{M}} [\mathcal{M} \models \phi = \neg\exists x\psi] \wedge \forall s \in M (\neg\psi(\underline{s}) \in A) \end{aligned}$$

Let  $\Gamma^{\mathcal{M}} : \mathcal{P}(M) \rightarrow \mathcal{P}(M)$  be the function defined

$$\Gamma^{\mathcal{M}}(A) = \{\phi \in \mathcal{M} \mid \Theta_{\mathcal{M}}(\phi, A)\} \quad (\Gamma)$$

Let us now define:

$$\begin{aligned} \Gamma^{\mathcal{M}}(0) &= \Gamma^{\mathcal{M}}(\emptyset) \\ \Gamma_{\alpha+1}^{\mathcal{M}} &= \Gamma^{\mathcal{M}}(\Gamma_{\alpha}^{\mathcal{M}}) \\ \Gamma_{\beta}^{\mathcal{M}} &= \bigcup_{\alpha < \beta} \Gamma_{\alpha}^{\mathcal{M}}, \text{ for } \beta \text{ - a limit ordinal} \end{aligned}$$

It can be checked that for some ordinal  $\alpha$  we must get  $\Gamma_{\alpha+1}^{\mathcal{M}} = \Gamma_{\alpha}^{\mathcal{M}}$ ; i.e.  $\Gamma_{\alpha}^{\mathcal{M}}$  is a fixpoint of  $\Gamma^{\mathcal{M}}$  (originally the argument has been presented in [30]; it is given also in [20]). In general, if  $A$  is any fixpoint of  $\Gamma^{\mathcal{M}}$ , then

$$(\mathcal{M}, A) \models \text{PT}^-$$

Let  $\alpha_{\mathcal{M}}$  denote the least ordinal  $\alpha$  such that  $\Gamma_{\alpha}^{\mathcal{M}}$  is a fixpoint of  $\Gamma^{\mathcal{M}}$ . Then  $\Gamma_{\alpha_{\mathcal{M}}}^{\mathcal{M}}$  is the least fixpoint of  $\Gamma^{\mathcal{M}}$ . Moreover

$$(\mathcal{M}, \Gamma(\alpha_{\mathcal{M}})) \models \text{PT}^- + \text{Cons}$$

Let us now elaborate on various properties of extensions for  $\text{PT}^-$  obtained in the above described way to better grasp both the advantages and limitations of this method. We have the following property isolated in the context of KF already by Cantini in [3]:

**Proposition 116.** Let  $\mathcal{M} \models \text{PA}$  and  $A \subseteq M$  be such that

$$(\mathcal{M}, A) \models \text{PT}^-$$

Then for  $B$  defined:

$$\phi \in B \iff \phi \in \text{Sent}_{\mathcal{L}_{\text{PA}}}^{\mathcal{M}} \wedge \neg\phi \notin A$$

we have also

$$(\mathcal{M}, B) \models \text{PT}^-$$

*Proof.* Let  $A, B$  be as above. Let us check the axioms for the conjunction and the negation of the universal quantifier. By definition and the fact that  $A$  is an extension for  $\text{PT}^-$ , we have (for arbitrary  $\psi_1, \psi_2$ , formula with at most one free variable  $\phi$  and variable  $v$ )

$$\begin{aligned} \psi_1 \wedge \psi_2 \in B &\iff \neg(\psi_1 \wedge \psi_2) \notin A \\ &\iff \neg\psi_1 \notin A \wedge \neg\psi_2 \notin A \\ &\iff \psi_1 \in B \wedge \psi_2 \in B \end{aligned}$$

and

$$\begin{aligned} \neg\forall v\phi \in B &\iff \neg\neg\forall v\phi \notin A \\ &\iff \exists a \in M \phi(\underline{a}) \notin A \\ &\iff \exists a \in M \neg\neg\phi(\underline{a}) \notin A \\ &\iff \exists a \in M \neg\phi(\underline{a}) \in B \end{aligned}$$

□

It transpires that there is a connection between  $\alpha_{\mathcal{M}}$  and the recursive saturation: in [5] it was shown that

**Lemma 117.** If  $\mathcal{M} \models \text{PA}$  is recursively saturated, then  $\alpha_{\mathcal{M}} = \omega$ .

The above reverses. To our best knowledge this is our original observation, however the technique is rather standard.

**Proposition 118.** If  $\alpha_{\mathcal{M}} = \omega$ , then  $\mathcal{M}$  is recursively saturated.

*Proof.* We prove the contraposition: suppose that a non-standard model  $\mathcal{M}$  is not recursively saturated. Let  $p(x)$  be a recursive type using parameters from  $\bar{a}$  which is omitted in  $\mathcal{M}$ . Let  $(\phi_i(x, \bar{y}))_i$  be an enumeration of formulae in  $p(x)$ . Without loss of generality, assume that  $\phi_0(x, \bar{y}) = (x = x)$ . Let

$$\psi_i(x, \bar{y}) = \bigwedge_{j < i} \phi_j(x, \bar{y}) \wedge \neg\phi_i(x, \bar{y})$$

Then every  $b \in M$  satisfies exactly one of  $\psi_i(x, \bar{a})$ . Now, for every  $n \in \omega$  we shall define formulae  $\theta_n(x)$ :

$$\begin{aligned} \theta_n^0(x) &= (x \neq x) \\ \theta_n^{k+1}(x, \bar{y}) &= \psi_{n-(k+1)}(x, \bar{y}) \vee \theta_n^k(x, \bar{y}) \\ \theta_n(x, \bar{y}) &= \theta_n^n(x, \bar{y}) \end{aligned}$$

Let us observe that the above construction can be arithmetised and therefore for some  $b \in M \setminus \mathbb{N}$  there exists a (code of) formula  $\theta_b(x, \bar{y})$ , which looks like

$$(\psi_0(x, \bar{y}) \vee (\psi_1(x, \bar{y}) \vee (\psi_2(x, \bar{y}) \vee (\dots (\psi_{b-1}(x, \bar{y}) \vee x \neq x) \dots)))$$

Then for each  $c \in M$  there exists  $n \in \omega$  such that  $\theta_b(\underline{c}, \bar{a}) \in \Gamma_n^{\mathcal{M}}$ , since each  $c$  satisfies some  $\psi_i(x, \bar{a})$  (because  $p(x)$  is omitted). But also for every  $i \in \omega$  there exists  $c \in M$  such that the least  $n$  for which we have  $\psi_n(c, \bar{a})$  is greater than  $i$ . Consequently, there is no  $k \in \omega$  for which

$$\theta_b(c, \bar{a}) \in \Gamma_k^{\mathcal{M}}$$

for every  $c \in M$ . In particular,  $\forall v \theta(v, \bar{x}) \notin \Gamma_\omega^{\mathcal{M}}$ , hence,  $\alpha_{\mathcal{M}} \neq \omega$ .  $\square$

Let us stress why the above theorem is relevant to our study: in every model,  $\Gamma_\omega$  is the sum of sets definable by a recursive sequence of arithmetical formulae. Most of conservativity results (including Theorem 225 given in Section 5) obtained for extensions of  $\text{PT}^-$  use models in which such an extension for  $\text{PT}^-$  can be found. Theorem 118 shows that our favourite techniques will not work for models which are not recursively saturated. In particular, estimating the model-theoretical strength of some extensions of  $\text{PT}^-$  (for example by showing that they admit models which are not recursively saturated) may require essentially new ideas.

We shall end this section with an easy observation on some non-conservative extensions of  $\text{PT}^-$  and  $\text{WPT}^-$ :

**Proposition 119.** *Both  $\text{PT}_1$  and  $\text{WPT}_1$  prove Tot and Cons. In particular, both theories are equal to  $\text{CT}_1$ .*

The argument uses formal induction on  $n$  in the formula

$$\forall \phi(\bar{v}) \left( \text{Compl}(\phi) \leq n \rightarrow \forall \alpha \in \text{Asn}(\phi) \left( \text{tot}(\phi[\alpha]) \wedge \text{cons}(\phi[\alpha]) \right) \right)$$

which is admissible in both  $\text{PT}_1$  and  $\text{WPT}_1$ , since the above is clearly a  $\Pi_1$  formula of  $\mathcal{L}_T$ . In Section 5 we shall demonstrate that Tot and Cons are provable already in  $\text{PT}_0$  and a natural extension of  $\text{WPT}_0$ .

### 3.4 Reflection Principles

The strength of basic compositional truth theories extended with various reflection principles involving the truth predicate provides particular interest for us in this dissertation. The Global Reflection ((REF)), one such principle, was already mentioned in the subsection devoted to strength of classically compositional truth theories. (REF) seems to be the most natural way of expressing that all theorems of PA are true, in the language  $\mathcal{L}_T$ . However, if one would like to justify such an assertion, another reflection principle seems to be involved: one would probably say that all theorems of PA are true, since all its axioms are and *truth is preserved by reasonings in First-Order Logic*; i.e. true premises lead to true conclusions. Let us isolate the latter principle:

**Definition 120.** The *First-Order Logic Closure Principle* is the following sentence of  $\mathcal{L}_T$ :

$$\forall \phi \ (\text{Pr}_\emptyset^T(\phi) \rightarrow T(\phi)),$$

where  $\text{Pr}_\emptyset^T(x)$  is the provability predicate for the  $\mathcal{L}_T$  theory  $T(x)$ .

At first glance, First-Order Logic Closure Principle seems to be neither implying, nor implied by the Global Reflection. The former says nothing about the truth of axioms of PA, while the latter is not a *closure* property: it does not seem to say that conclusions of true sentences are true. The latter can be seen as a *completeness* principle, saying that all sentences from a certain class are true. Let us isolate the completeness principle corresponding to First-Order Logic Closure Principle

**Definition 121.** The *First-Order Logic Completeness Principle* is the following sentence of  $\mathcal{L}_T$ :

$$\forall \phi \ (\text{Pr}_\emptyset(\phi) \rightarrow T(\phi)).$$

That, when added to  $\text{CT}^-$ , First-Order Logic Completeness Principle (and consequently, its closure counterpart) proves that all axioms of PA are true (and, consequently, Global Reflection) was first observed by Cezary Cieřliński:

**Theorem 122** (Cieřliński, [7]).  $\text{CT}^- + \forall \phi \ (\text{Pr}_\emptyset(\phi) \rightarrow T(\phi)) \vdash \forall \phi \ (\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi))$

Much later, the same author showed that, over  $\text{CT}^-$ , the two principles are in fact equivalent

**Theorem 123** (Cieřliński, [4]).  $\text{CT}^- + \forall \phi \ (\text{Pr}_\emptyset(\phi) \rightarrow T(\phi)) \vdash \forall \phi \ (\text{Pr}_\emptyset^T(\phi) \rightarrow T(\phi))$

We shall give a proof of this Theorem in Section 4. In Section 5, we shall show that over  $\text{PT}^-$  completeness principles, such as (REF), are strictly weaker than the closure ones. Let us isolate two more reflection principles obtained by weakening the logic we reason in:

**Definition 124.** *Propositional Logic Completeness Principle* is the following sentence of  $\mathcal{L}_T$ :

$$\forall \phi \ (\text{Pr}_{\text{CPC}}(\phi) \rightarrow T(\phi))$$

*Propositional Logic Closure Principle* is the following sentence of  $\mathcal{L}_T$ :

$$\forall \phi \ (\text{Pr}_{\text{CPC}}^T(\phi) \rightarrow T(\phi)),$$

where  $\text{Pr}_{\text{CPC}}^T$  is defined as in Definition 39.

Let us note that, similarly to  $\text{Pr}_\emptyset^T$ ,  $\text{Pr}_{\text{CPC}}^T$  expresses that true premises of reasoning in Classical Propositional Calculus lead to true results. In the next subsection, we shall give the context in which the two principles emerged and relate *Propositional Logic Closure Principle* to  $\Delta_0$  induction for the truth predicate. It is one of the consequences of the result from Section 4 that the above closure principle is, in fact, equivalent (over  $\text{CT}^-$ ) to previously introduced reflection principles.

Let us end this subsection by characterizing the set of  $\mathcal{L}_{\text{PA}}$  consequences of

$$\text{CT}^- + \forall \phi \ (\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi))$$

**Theorem 125.**  $\text{CT}^- + \forall\phi (\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi)) \vdash \mathcal{UR}^\omega$

*Proof.* By Theorem 123, it is sufficient to show that

$$\text{CT}^- + \forall\phi (\text{Pr}_{\text{PA}}^T(\phi) \rightarrow T(\phi)) \vdash \mathcal{UR}^\omega$$

where  $\text{Pr}_{\text{PA}}^T(x)$  is the provability predicate for the  $\Delta_1$   $\mathcal{L}_T$  theory  $x \in \text{PA} \vee T(x)$ . By induction on  $n$  we show that

$$\text{CT}^- + \forall\phi (\text{Pr}_{\text{PA}}^T(\phi) \rightarrow T(\phi)) \vdash \forall\phi (\text{Pr}_{\mathcal{UR}^n}(\phi) \rightarrow T(\phi))$$

which will suffice, since for arbitrary formula  $\phi(x_0, \dots, x_n)$

$$\text{PA} \vdash \forall x_0, \dots, x_n (\text{Sent}_{\mathcal{L}_{\text{PA}}}(\ulcorner \phi(\underline{x}_0, \dots, \underline{x}_n) \urcorner))$$

and

$$\text{CT}^- \vdash \forall x_0, \dots, x_n (T(\ulcorner \phi(\underline{x}_0, \dots, \underline{x}_n) \urcorner) \equiv \phi(x_0, \dots, x_n)).$$

The base step for  $n = 0$  is obvious, since  $\mathcal{UR}^0 = \text{PA}$ . Assume that the above holds for  $n$ . We reason in  $\text{CT}^- + \forall\phi (\text{Pr}_{\text{PA}}^T(\phi) \rightarrow T(\phi))$ . Since the set of true sentences is closed under First-Order Logic, it is sufficient to show that for every sentence  $\phi \in \mathcal{L}_{\text{PA}}$ , we have

$$T(\text{Pr}_{\mathcal{UR}^n}(\phi) \rightarrow \phi)$$

So fix arbitrary sentence  $\phi \in \mathcal{L}_{\text{PA}}$ . By our induction assumption we have

$$\text{Pr}_{\mathcal{UR}^n}(\phi) \rightarrow T(\phi)$$

Using the compositional axiom for  $T$  we obtain

$$T(\text{Pr}_{\mathcal{UR}^n}(\phi) \rightarrow \phi)$$

which ends the proof. □

Let us now show that  $\mathcal{UR}^\omega$  suffices for characterizing the set of arithmetical consequences of  $\text{CT}^- + \forall\phi (\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi))$ .

**Theorem 126** (Kotlarski, [28]). *For every sentence  $\phi$  of  $\mathcal{L}_{\text{PA}}$ , if  $\text{CT}^- + \forall\phi (\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi)) \vdash \phi$ , then  $\mathcal{UR}^\omega(\text{PA}) \vdash \phi$ .*

In such cases we will also say, that  $\text{CT}^- + \forall\phi (\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi))$  is *proof-theoretically conservative* over  $\mathcal{UR}^\omega(\text{PA})$  (stretching the scope of Definition 81 a little). Being more precise, Kotlarski proved the above conservativity result for the above theory of truth and a different theory that we shall define right now. However, the link between  $\mathcal{UR}^\omega(\text{PA})$  and Kotlarski's theory is quite clear, so we do not consider this modification an essentially new contribution of this dissertation.

**Definition 127.** We shall define a sequence of  $\mathcal{L}_{\text{PA}}$  formulae  $\{\Xi_n(x)\}_{n \in \omega}$ :

$$\begin{aligned}\Xi_0(x) &:= \text{Pr}_{\text{PA}}(x) \\ \Xi'_{n+1}(x) &:= x \in \text{PA} \vee \exists \psi(v) ((x = \forall v \psi(v)) \vee \forall y \Xi_n(\psi(\underline{y}))) \\ \Xi_{n+1}(x) &:= \text{Pr}_{\Xi'_{n+1}}(x)\end{aligned}$$

Now define  $\Xi(\text{PA}) = \text{PA} \cup \{\neg \Xi_n(\ulcorner 0 = 1 \urcorner) \mid n \in \omega\}$ .

**Remark 128.**

1. For every  $n$ ,  $\Xi_n(x)$  is a formula of  $\mathcal{L}_{\text{PA}}$  of  $\Sigma_{2n+1}$  class.
2. By definition  $\text{Con}(\Xi'_n) = \neg \Xi_{n+1}(\ulcorner 0 = 1 \urcorner)$ .

*Proof of Theorem 126.* Theorem 5.1 in [28] shows that  $\text{CT}^- + \forall \phi (\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi))$  is proof-theoretically conservative over  $\Xi(\text{PA})$ , so it suffices to show that

$$\mathcal{UR}^\omega(\text{PA}) \vdash \Xi(\text{PA}).$$

To this end, by induction on  $n$  we show that for all  $n$

$$\text{PA} \vdash \forall \phi (\Xi_n(\phi) \rightarrow \text{Pr}_{\text{Tr}_{\mathcal{UR}^n \upharpoonright_{2n+1}}}(\phi)), \quad (*)$$

which obviously will suffice, since for each  $n$ , the consistency of  $\text{Tr}_{\mathcal{UR}^n \upharpoonright_{2n+1}}$  is provable in  $\mathcal{UR}^{n+1}$  (Proposition 72,  $\text{Tr}_{\mathcal{UR}^n \upharpoonright_n}$  is as in Definition 69). The base step for  $n = 0$  follows by definitions of respective theories. Now, assume that  $*$  holds for an  $n$ . We reason in PA. Fix  $\phi$  and assume  $\Xi_{n+1}(\phi)$ . Then there exists  $\psi(v)$  such that

$$\begin{aligned}\Xi_n(\forall v \psi(v) \rightarrow \phi) \\ \forall x \Xi_n(\psi(\underline{x}))\end{aligned}$$

By induction assumption  $*$  for  $n$ , we know that

$$\begin{aligned}\text{Pr}_{\text{Tr}_{\mathcal{UR}^n \upharpoonright_{2n+1}}}(\forall v \psi(v) \rightarrow \phi) \\ \forall x \text{Pr}_{\text{Tr}_{\mathcal{UR}^n \upharpoonright_{2n+1}}}(\psi(\underline{x}))\end{aligned}$$

Since  $\forall x \text{Pr}_{\text{Tr}_{\mathcal{UR}^n \upharpoonright_{2n+1}}}(\psi(\underline{x}))$  is a true  $\Sigma_{2n+3}$  sentence and  $\text{Tr}_{\mathcal{UR}^n \upharpoonright_{2n+1}} \subseteq_{2n+4} \text{Tr}_{\mathcal{UR}^{n+1} \upharpoonright_{2n+3}}$ , then

$$\text{Pr}_{\text{Tr}_{\mathcal{UR}^{n+1} \upharpoonright_{2n+3}}}(\forall v \psi(v) \rightarrow \phi) \quad (**)$$

$$\text{Pr}_{\text{Tr}_{\mathcal{UR}^{n+1} \upharpoonright_{2n+3}}}(\ulcorner \forall v \text{Pr}_{\text{Tr}_{\mathcal{UR}^n \upharpoonright_{2n+1}}}(\psi(\underline{v})) \urcorner)$$

By the reflection axioms for  $\text{Tr}_{\mathcal{UR}^n \upharpoonright_{2n+1}}$  in  $\mathcal{UR}^{n+1}$  (Lemma 71) we obtain

$$\text{Pr}_{\text{Tr}_{\mathcal{UR}^{n+1} \upharpoonright_{2n+3}}}(\forall v \psi(v))$$

Putting this together with (\*\*), we obtain  $\text{Pr}_{\text{Tr}_{\mathcal{UR}^{n+1} \upharpoonright_{2n+3}}}(\phi)$  as wanted.  $\square$

## 3.5 Additional Axioms

In addition to the above axioms, we will often consider two more different principles (together with their variations). The first one is a straightforward generalisation of the axiom for disjunction admissible in compositional theories. Let us first define it and then justify its importance.

**Definition 129** (PA). Let  $\phi_0, \dots, \phi_x$  be a parametrized family of formulae (of length  $x + 1$ ). By

$$\bigvee_{i \leq x} \phi_i$$

we denote the disjunction of  $\phi_0, \dots, \phi_x$ , i.e. the following  $\mathcal{L}_{PA}$  formula

$$\begin{aligned} \psi_0 &= (\phi_0) \\ \psi_{n+1} &= (\phi_{n+1}) \vee \psi_n \end{aligned}$$

and  $\bigvee_{i < x} \phi_i = \psi_x$ . Moreover if  $c$  is any set of sentences then we will also write

$$\bigvee_{\phi \in c} \phi$$

to denote

$$\bigvee_{i < x} \phi_i$$

where  $x$  is the cardinality of  $c$  and  $\phi_0, \dots, \phi_{x-1}$  is the growing enumeration of formulae from  $c$ . If  $\phi_0, \dots, \phi_x$  is any parametrized family of formulae, then as expected we treat  $\bigwedge_{i \leq x} \phi_i$  as the abbreviation of

$$\neg \bigvee_{i \leq x} \neg \phi_i$$

**Definition 130** (Disjunctive Correctness for the Satisfaction Relation). The axiom of *Disjunctive Correctness for the Satisfaction Relation* is the following sentence of  $\mathcal{L}_S$ :

$$\forall c \left( \text{SetSent}(c) \rightarrow \forall \alpha \in \text{Asn}(c) \left( S \left( \bigvee_{\phi \in c} \phi, \alpha \right) \equiv \exists \phi \in c S(\phi, \alpha \upharpoonright) \right) \right) \quad (\text{DC}(S))$$

**Definition 131** (Disjunctive Correctness). The axiom of *Disjunctive Correctness* is the following sentence of  $\mathcal{L}_T$ :

$$\forall c \left( \text{SetSent}(c) \rightarrow \left( T \left( \bigvee_{\phi \in c} \phi \right) \equiv \exists \phi \in c T(\phi) \right) \right) \quad (\text{GV})$$

The interest in both principles was initiated by a corollary to Kotlarski, Krajewski and Lachlan proof of Theorem 108, noticed already in the original paper. It stated that  $\text{DC}(S)$  is not provable in  $\text{CS}^-$  and the proof gave a construction of a model  $\mathcal{M} \models \text{CS}^-$ , such that for some non-standard number  $a$  the pair

$$\left( \bigvee_{i < a} 0 = 1, \varepsilon \right)$$

is in the extension of  $S$  in  $\mathcal{M}$ . Hence, a natural question can be asked: how strong are theories that do not admit pathological models of this kind? Much later in [6] the strength of Propositional Logic Closure Principle (as in Definition 124) was studied precisely with this motivation. It was shown that

**Theorem 132** (Cieśliński, [6]).  $CT_0$  and  $CT^- + \forall\phi(\text{Pr}_{\text{CPC}}^T(\phi) \rightarrow T(\phi))$  are deductively equivalent.

The above theorem, at the time it was proven, was believed to show (compare our discussion in Subsection 3.3.3) that  $CT^-$  with the Propositional Logic Closure Principle added is a very strong theory. It is worth emphasizing that, over  $CT^-$ , the above principle is equivalent to the conjunction of DC and the Propositional Logic Completeness Principle (as in Definition 124). Unfortunately, the strength of both principles separately has yet to be determined.

Since in  $PT^-$  we do not have the global axiom for the negation, then together with the usual axiom  $G\forall$  (see Definition 131), we have to add the axiom saying when the negation of a generalised disjunction is true: i.e.

$$\forall c \left( \text{SetSent}(c) \rightarrow \left( T(\neg \bigvee_{\phi \in c} \phi) \equiv \forall \phi \in c T(\neg \phi) \right) \right) \quad (G\neg\forall)$$

Let us now introduce the analogue of the Disjunctive Correctness matching the intuitions behind the Weak Kleene Logic. As we will not consider disjunctively correct extensions of  $WPS^-$ , we introduce this principle only for  $\mathcal{L}_T$ :

**Definition 133.** The *Weak Kleene Disjunctive Correctness Principle* is the following sentence of  $\mathcal{L}_T$

$$\forall c \left( \text{SetSent}(c) \rightarrow \left( T\left(\bigvee_{\phi \in c} \phi\right) \equiv \left( (\forall \phi \in c \text{tot}(\phi)) \wedge (\exists \phi \in c T(\phi)) \right) \right) \right) \quad (G\forall_{wk})$$

Let us define the disjunctively correct extensions of our theories:

**Definition 134.**

1.  $CT^- + DC$  is the theory  $CT^- + G\forall$
2.  $PT^- + DC$  is the theory  $PT^- + G\forall + G\neg\forall$
3.  $WPT^- + DC$  is the theory  $WPT^- + G\forall_{wk} + G\neg\forall$ .

For completeness, let us state the following proposition:

**Proposition 135.** Both  $PT^- + DC$  and  $WPT^- + DC$  (and, consequently,  $CT^-$ ) prove the following sentence

$$\forall c \left( \text{SetSent}(c) \rightarrow \left( T\left(\bigwedge_{\phi \in c} \phi\right) \equiv \forall \phi \in c T(\phi) \right) \right)$$

Moreover  $PT^- + DC$  proves

$$\forall c \left( \text{SetSent}(c) \rightarrow \left( T(\neg \bigwedge_{\phi \in c} \phi) \equiv \exists \phi \in c T(\neg \phi) \right) \right)$$

and  $WPT^- + DC$  proves

$$\forall c \left( \text{SetSent}(c) \rightarrow \left( T(\neg \bigwedge_{\phi \in c} \phi) \equiv \left( (\forall \phi \in c \text{tot}(\phi)) \wedge (\exists \phi \in c T(\neg \phi)) \right) \right) \right)$$

The next additional axiom we will consider is the analogue of the Induction Axiom known from Second Order Arithmetic (it is an axiom of e.g.  $ACA_0$ ).

**Definition 136.** The *Internal Induction Axiom* is the following sentence of  $\mathcal{L}_T$ :

$$\forall v \forall \phi(v) \left( (T(\phi(\underline{0})) \wedge \forall x (T(\phi(\underline{x})) \rightarrow T(\phi(\underline{x+1})))) \rightarrow \forall x T(\phi(\underline{x})) \right) \quad (\text{INT})$$

In the context of non-classically compositional theories, it makes perfect sense to consider restricted versions of internal induction. One of such version was introduced in [12] and later studied in [13] and [15]:

**Definition 137.** The *Internal Induction Axiom for Total Formulae* is the following sentence of  $\mathcal{L}_T$ :

$$\forall v \forall \phi(v) \left( \text{tot}(\phi) \rightarrow (T(\phi(\underline{0})) \wedge \forall x (T(\phi(\underline{x})) \rightarrow T(\phi(\underline{x+1})))) \rightarrow \forall x T(\phi(\underline{x})) \right) \quad (\text{INT}(\text{tot}))$$

Let  $\text{PT}^- + \text{INT}(\text{tot})$ , be  $\text{PT}^-$  with the above sentence added. The first study we know that investigates this type of axiom is [2], where it was studied in the context of untyped compositional truth theory KF. The problem of its strength over typed theories of truth was raised in [12] (and later in [13],[14],[15]) where it was claimed that  $\text{PT}^- + \text{INT}(\text{tot})$  is an example of an axiomatic theory of truth which is:

1. non-interpretable in PA (for the definition see the original paper [12]) but
2. model-theoretically conservative over it.

In order to see why the first one holds let us start by noticing that all the compositional truth theories we consider, when extended with INT are finitely axiomatisable. More concretely:

**Proposition 138.** *Let  $\phi$  be the conjunction of compositional axioms of either  $\text{PT}^-$ , or  $\text{WPT}^-$ , or  $\text{CT}^-$ . Then the following theories are deductively equivalent:*

1.  $\text{PA} + \phi + \text{INT}$ ,
2.  $\text{PA} + \phi + \text{INT}(\text{tot})$ ,
3.  $I\Sigma_1 + \phi + \text{INT}$ ,
4.  $I\Sigma_1 + \phi + \text{INT}(\text{tot})$ .

The proof has essentially been given in [12] and [13] in the context of  $\text{PT}^-$ , but analogous reasoning works for the rest of theories considered. Its basic idea is that INT (and INT(tot)) expresses all the instantiations of induction scheme in a single sentence, hence, to axiomatise  $\text{Th}^- + \text{INT}$  we only need a finitely axiomatised basic theory of syntax (like  $I\Sigma_1$ ), finitely many compositional axioms of Th and INT. What is more, this proof can be arithmetised, so PA proves that every induction axiom can be deduced from the chosen axiom of the respective truth theory. For the details, we refer the Reader to the literature.

However,  $\text{PT}^- + \text{INT}(\text{tot})$ , contrary to what was claimed in [12], fails to realise the second aim for which it was designed: it was proven to be model-theoretically stronger than PA.

**Theorem 139** (Wcisło,[5]).  $PT^- + INT(\text{tot})$  is not model-theoretically conservative over PA.

We will continue discussing the extensions of  $WPT^-$  and  $PT^-$  with INT in Section 5.2. One of the original results of this section is that  $WPT^- + DC + INT$  is model-theoretically conservative over PA and that its Strong-Kleene counterpart,  $PT^- + DC + INT$ , is the same theory as  $CT_0$ . For now, let us indicate, that adding INT does not always proof-theoretically strengthen the theory: we state it for  $CT^-$ , since the rest of theories are its subtheories:

**Theorem 140** (Kotlarski-Krajewski-Lachlan [29], Enayat-Visser, Leigh[31]).  $CT^- + INT$  is syntactically conservative over PA.

As in the case of Theorem 108, it was originally obtained in [29] for  $CS^-$ . Arguments given in [9] and [31] provide alternative proof methods.

## 4. CLASSICALLY COMPOSITIONAL TRUTH THEORIES

### 4.1 Classical Compositional Truth with the $\Delta_0$ -induction

We shall estimate the strength of  $CS_0$  and  $CT_0$ . Most importantly, we shall prove the following theorem

**Theorem 141.**  $CT_0$  proves the Global Reflection Principle.

In fact, we shall demonstrate a strengthening of the above: we will prove

**Theorem 142.** The following sentence is provable in  $CS_0$ :

$$\forall\phi \text{ Pr}_{\text{PA}}(\phi) \rightarrow \forall\alpha \in \text{Asn}(\phi) S(\phi, \alpha)$$

The latter theorem is the exact analogue of the former in the context of the satisfaction predicate. The most important conclusion that is to be drawn from the main result of this thesis is that induction for the  $\Delta_0$ -formulae of the extended language suffices to prove that the satisfaction (truth) predicate is extremely well-behaved. The key point to do this is to establish that a kind of induction on the build-up of formulae is admissible in  $CS_0$ . The non-obviousness of this result has already been discussed in Section 3.3.3. The most important lemma on our way is to show that  $CS_0$  proves the generalised axiom for commutation of the satisfaction predicate with blocks of universal quantifiers. It is the content of the theorem below, but it will be useful to introduce a piece of notation first:

**Definition 143 (PA).** 1.  $\sigma, \tau$  range over (internally) finite sequences (i.e. functions whose domain is (internally) finite and closed downwards) of variables.

2. If  $\phi$  is a sentence and  $\sigma$  is a sequence of variables, then  $\text{ucl}(\sigma, \phi)$  denotes the universal closure of  $\phi$  w.r.t.  $\sigma$ , i.e. a sentence of the form

$$\forall\sigma(0) \dots \forall\sigma(a)\phi$$

where  $a = \max(\text{dom}(\sigma))$ . Note that  $\psi = \text{ucl}(\sigma, \phi)$  can be expressed with a  $\Delta_0$  formula (we examine the structure of  $\psi$ ). We do not demand  $\sigma$  to be injective, neither to enumerate only free variables of  $\phi$ . Similarly, let  $\text{bucl}(\sigma, \phi, t)$  denote the bounded universal closure of  $\phi$  w.r.t.  $\sigma$ , i.e. the sentence

$$\forall\sigma(0) < t \dots \forall\sigma(a) < t\phi$$

3. If  $\sigma$  is a sequence and  $\alpha, \beta$  are assignments, then  $\alpha \sim_\sigma \beta$  means that  $\alpha$  differs from  $\beta$  at most on the values assigned to variables occurring in  $\sigma$  and is defined on the variables occurring in  $\sigma$ . We note that similarly to  $\sim_v$ , this relation is also not an equivalence relation.

4. If  $\alpha$  is an assignment,  $\sigma$  is a sequence of variables and  $b$  is a number, then

$$\alpha[\sigma \mapsto b]$$

denotes the unique assignment which assigns  $b$  to all the variables from sequence  $\sigma$  and does not differ from  $\alpha$  on the rest of variables. We note that  $\beta = \alpha[\sigma \mapsto b]$  can be expressed as a  $\Delta_0$  formula.

5.  $\preceq$  denotes the following relation between assignments (this is a slight variation of standard product ordering):

$$\alpha \preceq \beta \iff \text{dom}(\alpha) \subseteq \text{dom}(\beta) \wedge \forall z \in \text{dom}(\alpha) (\alpha(z) \leq \beta(z))$$

Note that  $\preceq$  is arithmetically definable with  $\Delta_0$ -formula and that

$$\text{PA} \vdash \forall \alpha \forall \beta (\alpha \preceq \beta \rightarrow \alpha \leq \beta)$$

**Theorem 144.** *The following sentence is provable in  $\text{CS}_0$*

$$\forall \phi \forall \sigma \forall \alpha \in \text{Asn}(\text{ucl}(\sigma, \phi)) \left( S(\text{ucl}(\sigma, \phi), \alpha) \equiv \forall \beta \sim_{\sigma} \alpha \ S(\phi, \beta \upharpoonright_{\phi}) \right)$$

It is convenient to split the proof into three lemmata. The first states that the generalised axiom for commutation with blocks of universal quantifiers holds for bounded quantifiers:

**Lemma 145.** *The following sentence is provable in  $\text{CS}_0$ :*

$$\forall \phi \forall \sigma \forall v \in \text{Var} \setminus \text{Var}(\text{ucl}(\sigma, \phi)) \forall \alpha \in \text{Asn}(\text{bucl}(\sigma, \phi, v)) \left( S(\text{bucl}(\sigma, \phi, v), \alpha) \equiv \forall \beta \left[ (\beta \sim_{\sigma} \alpha \wedge \beta \preceq \alpha[\sigma \mapsto (\alpha(v) - 1)]) \rightarrow S(\phi, \beta \upharpoonright_{\phi}) \right] \right)$$

*Proof.* We work in  $\text{CS}_0$ . Let us fix  $\phi, v, \sigma, \alpha$ , let  $a = \max(\text{dom}(\sigma))$  and  $b = \alpha(v) - 1$ . Moreover let

$$\gamma := \alpha[\sigma \mapsto b]$$

We have to show that

$$S(\text{bucl}(\sigma, \phi, v), \alpha) \equiv \forall \beta \left[ (\beta \sim_{\sigma} \alpha \wedge \beta \preceq \gamma) \rightarrow S(\phi, \beta \upharpoonright_{\phi}) \right]$$

Let  $\sigma \upharpoonright_n$  denote *first*  $n$  elements of sequence  $\sigma(0), \dots, \sigma(a)$  i.e.

$$\sigma(0), \dots, \sigma(n-1)$$

( $\sigma \upharpoonright_0$  is the empty sequence,  $\sigma \upharpoonright_{a+1} = \sigma$ .) Dually, let  $\sigma \downarrow^n$  denote the sequence consisting of *last*  $n$  elements of sequence  $\sigma$ , i.e.

$$\sigma(a - (n-1)), \sigma(a - (n-2)), \dots, \sigma(a)$$

(similarly  $\sigma \downarrow^0$  is the empty sequence,  $\sigma \downarrow^{a+1} = \sigma$ ). Similarly  $\gamma \upharpoonright_n$  is

$$\alpha[\sigma \upharpoonright_n \mapsto b].$$

( $\gamma \upharpoonright_n$  is really a  $\Delta_0$ -formula  $\gamma(x, n)$ . All the quantifiers are bounded by  $\gamma$  which is our parameter.) Let us note that  $\gamma \upharpoonright_0 = \alpha$  and  $\gamma \upharpoonright_{a+1} = \gamma$ . We write  $\phi_n$  for

$$\text{bucl}(\sigma \upharpoonright^n, \phi, v) \quad (= \forall \sigma(a - n + 1) < v \dots \forall \sigma(a) < v \phi).$$

( $\phi_0 = \phi$ ). By induction on  $n$  from 0 up to  $a + 1$  we show that for every  $0 \leq n \leq a + 1$

$$\begin{aligned} \forall \beta \left[ \beta \sim_{\sigma \upharpoonright_{a+1-n}} \alpha \wedge \beta \preceq \gamma \upharpoonright_{a+1-n} \longrightarrow S(\phi_n, \beta \upharpoonright_{\phi_n}) \right] \equiv \\ \forall \beta \left( \beta \sim_{\sigma} \alpha \wedge \beta \preceq \gamma \longrightarrow S(\phi, \beta \upharpoonright_{\phi}) \right) \quad (\text{IND}) \end{aligned}$$

(we keep adding bounded quantifiers starting from that which is closest to  $\phi$ .) Let us note that the left-hand side of the above for  $n = a + 1$  is equivalent to

$$\forall \beta \left( \beta \sim_{\sigma \upharpoonright_0} \alpha \wedge \beta \preceq \gamma \upharpoonright_0 \longrightarrow S(\phi_{a+1}, \beta \upharpoonright_{\phi_{a+1}}) \right)$$

and hence to

$$S(\text{bucl}(\phi, \sigma, v), \alpha)$$

We have to justify that (IND) satisfies induction axiom. All quantifiers binding  $\beta$  are bounded by  $\gamma \upharpoonright_n$  which in turn can be bounded by  $\gamma$  (for every  $n$ ). Moreover all formulae occurring in  $S$  predicate are of the form  $\phi_n$ , for  $n \leq a + 1$ . The greatest such formula is  $\phi_{a+1}$ , hence if  $d = \langle \phi_{a+1}, \gamma \rangle + 1$ , then we have

$$\text{IND}(n) \equiv \text{IND}[S(x, y) \wedge \langle x, y \rangle < d / S(x, y)](n)$$

for all  $n$ . The application of Remark 57 completes our justification. Let us verify the base and the induction step. Observe that the former is trivial: for  $n = 0$  we have the same formula on both sides of the equivalence sign. We shall present the induction step for  $n = 1$  first and then the proof in full generality: hopefully it will help the Reader to keep track of progress in our proof. We start from the left-hand side of IND:

$$\forall \beta \left[ \beta \sim_{\sigma \upharpoonright_a} \alpha \wedge \beta \preceq \gamma \upharpoonright_a \longrightarrow S(\phi_1, \beta \upharpoonright_{\phi_1}) \right]$$

By the definition of  $\phi_1$  and axioms of  $\text{CS}^-$  the above is equivalent to

$$\forall \beta \left[ \beta \sim_{\sigma \upharpoonright_a} \alpha \wedge \beta \preceq \gamma \upharpoonright_a \longrightarrow \forall \zeta \left[ \zeta \sim_{\sigma(a)} \beta \wedge (\sigma(a))_{\zeta}^{\circ} < (v)_{\zeta}^{\circ} \rightarrow S(\phi, \zeta \upharpoonright_{\phi}) \right] \right]$$

Now, by the properties of the assignment function, provable in PA and the fact that  $\zeta(v) = \beta(v) = \alpha(v) = b$ , since  $v \notin \text{Var}(\text{ucl}(\sigma, \phi))$  we have

$$\forall \beta \left( \beta \sim_{\sigma \upharpoonright_a} \alpha \wedge \beta \preceq \gamma \upharpoonright_a \longrightarrow \forall \zeta \left[ \zeta \sim_{\sigma(a)} \beta \wedge \zeta(a) < b \rightarrow S(\phi, \zeta \upharpoonright_{\phi}) \right] \right)$$

By logic the above is equivalent to

$$\forall \beta \forall \zeta \left( (\beta \sim_{\sigma \upharpoonright_a} \alpha \wedge \beta \preceq \gamma \upharpoonright_a \wedge \zeta \sim_{\sigma(a)} \beta \wedge \zeta(a) < b) \longrightarrow S(\phi, \zeta \upharpoonright_{\phi}) \right)$$

Now, because if  $\beta \sim_{\sigma \upharpoonright_a} \alpha$  and  $\zeta \sim_{\sigma(a)} \beta$ , then  $\zeta \sim_{\sigma \upharpoonright_{a+1}} \alpha$  and if  $\beta \preceq \gamma \upharpoonright_a$  and  $\zeta \sim_{\sigma(a)} \beta$  and  $\zeta(a) < b$ , then  $\zeta \preceq \gamma \upharpoonright_{a+1} = \gamma$ . So the above is equivalent to

$$\forall \beta \left( \beta \sim_{\sigma} \alpha \wedge \beta \preceq \gamma \longrightarrow S(\phi, \beta \upharpoonright_{\phi}) \right)$$

which is precisely what we wanted.

The proof of the induction step involves the same justifications, so we state it line by line (we start with the left-hand side of (IND) for  $n + 1$  and the following are equivalent)

1.  $\forall \beta \left[ \beta \sim_{\sigma \upharpoonright_{a+1-(n+1)}} \alpha \wedge \beta \preceq \gamma \upharpoonright_{a+1-(n+1)} \longrightarrow S(\phi_{n+1}, \beta \upharpoonright_{\phi_{n+1}}) \right]$
2.  $\forall \beta \left[ \beta \sim_{\sigma \upharpoonright_{a+1-(n+1)}} \alpha \wedge \beta \preceq \gamma \upharpoonright_{a+1-(n+1)} \longrightarrow \right.$   
 $\left. \longrightarrow \forall \zeta \left[ \zeta \sim_{\sigma(a+1-n)} \beta \wedge (\sigma(a+1-n))_{\zeta}^{\circ} < (y)_{\zeta}^{\circ} \rightarrow S(\phi_n, \zeta \upharpoonright_{\phi_n}) \right] \right]$
3.  $\forall \beta \left( \beta \sim_{\sigma \upharpoonright_{a+1-(n+1)}} \alpha \wedge \beta \preceq \gamma \upharpoonright_{a+1-(n+1)} \longrightarrow \right.$   
 $\left. \longrightarrow \forall \zeta \left[ \zeta \sim_{\sigma(a+1-n)} \beta \wedge \zeta(a+1-n) < b \rightarrow S(\phi_n, \zeta \upharpoonright_{\phi_n}) \right] \right)$
4.  $\forall \beta \forall \zeta \left( (\beta \sim_{\sigma \upharpoonright_{a-n}} \alpha \wedge \beta \preceq \gamma \upharpoonright_{a-n} \wedge \zeta \sim_{\sigma(a+1-n)} \beta \wedge \zeta(a+1-n) < b) \longrightarrow \right.$   
 $\left. \longrightarrow S(\phi_n, \zeta \upharpoonright_{\phi_n}) \right)$
5.  $\forall \beta \left( \beta \sim_{\sigma \upharpoonright_{a+1-n}} \alpha \wedge \beta \preceq \gamma \upharpoonright_{a+1-n} \longrightarrow S(\phi_n, \beta \upharpoonright_{\phi_n}) \right)$

By inductive assumption the last sentence in this sequence is equivalent to

$$\forall \beta \left( \beta \sim_{\sigma} \alpha \wedge \beta \preceq \gamma \longrightarrow S(\phi, \beta \upharpoonright_{\phi}) \right)$$

which ends the proof. □

The next lemma is rather a straightforward application of Lemma 145.

**Lemma 146.** *The following sentence is provable in  $CS_0$*

$$\forall \phi(\bar{w}) \forall \sigma \forall v \in \text{Var} \setminus \text{Var}(\text{ucl}(\sigma, \phi)) \forall \alpha \in \text{Asn}(\text{ucl}(\sigma, \phi))$$

$$\left( S(\forall v \text{bucl}(\sigma, \phi, v), \alpha) \equiv \forall \beta \sim_{\sigma} \alpha \ S(\phi, \beta \upharpoonright_{\phi}) \right)$$

*Proof.* We work in  $CS_0$  and let us fix  $\phi, v, \sigma, \alpha$  as in our assumptions. Then, by the axiom for universal quantifier, we have

$$S(\forall v \text{bucl}(\sigma, \phi, v), \alpha) \equiv \left( \forall \gamma \sim_v \alpha \ S(\text{bucl}(\sigma, \phi, v), \gamma \upharpoonright_{\cdot}) \right)$$

Let us fix an arbitrary  $\gamma$  as in the above formula. Then, by Lemma 145, the above is equivalent to

$$\forall\beta \sim_{\sigma} \gamma \left( \beta \preceq \gamma[\sigma \mapsto \gamma(v)] \longrightarrow S(\phi, \beta \upharpoonright_{\phi}) \right) \quad (4.1)$$

Now by the fact that  $\gamma$  was arbitrary and differed from  $\alpha$  only on the value assigned to the variable  $v$  (which occurs in neither of  $\phi, \sigma$ ) the above is equivalent to

$$\forall\beta \sim_{\sigma} \alpha \ S(\phi, \beta \upharpoonright_{\phi}) \quad (4.2)$$

which ends the proof.  $\square$

The last lemma closes the proof of Theorem 144. The fact that we can prove it by using basically the same technique as in Lemma 145 was first observed by Bartosz Wcisło. However, the strategy of proving Theorem 144 via Lemmata 145, 146, 147 was our original idea and we (Bartosz Wcisło and the author) both agreed that the crucial insight was to isolate Lemma 145.

**Lemma 147.** *The following sentence is provable in  $CS_0$ :*

$$\forall\phi\forall\sigma\forall v \in \text{Var} \setminus \text{Var}(\text{ucl}(\sigma, \phi))\forall\alpha \in \text{Asn}(\text{ucl}(\sigma, \phi)) \left( S(\text{ucl}(\sigma, \phi), \alpha) \equiv S(\forall v \text{bucl}(\sigma, \phi, v), \alpha) \right)$$

*Proof.* We work in  $CS_0$ . Let us fix  $\phi, \sigma, v$  and  $\alpha$ , as in our assumptions and let  $a = \max(\text{dom}(\sigma))$ . The method is fully analogous to that used in the proof of Lemma 145: we inductively add unbounded universal quantifiers to  $\phi$  one in each step. More precisely, let  $\sigma \upharpoonright_n$  and  $\sigma \upharpoonright^n$  be as in the proof of Lemma 145. Let  $\phi_n$  now denote  $\text{ucl}(\phi, \sigma \upharpoonright^n)$ , i.e.<sup>1</sup>

$$\forall\sigma(a - n + 1) \dots \forall\sigma(a) \ \phi$$

By induction on  $n$  up to  $a + 1$  we show that

$$S(\forall v \text{bucl}(\sigma \upharpoonright_n, \phi_{a-n+1}, v), \alpha) \equiv S(\text{ucl}(\sigma, \phi), \alpha)$$

Let us note that  $\text{bucl}(\sigma \upharpoonright_n, \phi_{a-n+1}, v)$  is simply the result of bounding  $n$  first quantifiers in  $\text{ucl}(\sigma, \phi)$  with  $v$ , i.e.

$$\forall\sigma(0) < y \dots \forall\sigma(n - 1) < y \forall\sigma(n) \dots \forall\sigma(a) \phi$$

Observe also that for all  $n \leq a + 1$ ,

$$\langle \forall v \text{bucl}(\sigma \upharpoonright_n, \phi_{a-n+1}, v), \alpha \rangle < \langle \forall y \text{bucl}(\sigma, \phi, v), \alpha \rangle$$

hence, by Remark 57 we are allowed to use induction for the above formula. The base step is once again trivial, since it reduces to

$$S(\forall y \text{ucl}(\sigma, \phi), \alpha) \equiv S(\text{ucl}(\sigma, \phi), \alpha).$$

<sup>1</sup> It might be easier to understand what  $\text{ucl}(\phi, \sigma \upharpoonright^n)$  is when seeing it as generated by the following recursive procedure

$$\begin{aligned} \text{ucl}(\phi, \sigma \upharpoonright^0) &= \phi \\ \text{ucl}(\phi, \sigma \upharpoonright^{n+1}) &= \forall\sigma(a - n) \text{ucl}(\phi, \sigma \upharpoonright^n) \end{aligned}$$

Which is true by the axiom for universal quantifier in  $CS^-$  and the fact that  $y$  does not occur in  $\text{ucl}(\sigma, \phi)$ . In order to prove the inductive step it is sufficient to show that for every  $n \leq a + 1$

$$S(\forall y \text{bucl}(\sigma \upharpoonright_n, \phi_{a-n+1}, v), \alpha) \equiv S(\text{bucl}(\sigma \upharpoonright_{n+1}, \phi_{a-n}, v), \alpha)$$

So let us fix  $n \leq a + 1$  and write the following chain of equivalences

$$S(\forall v \text{bucl}(\sigma \upharpoonright_n, \phi_{a-n+1}, v), \alpha) \tag{4.3}$$

$$\forall \beta \sim_v \alpha \ S(\text{bucl}(\sigma \upharpoonright_n, \phi_{a-n+1}, v), \beta \upharpoonright.) \tag{4.4}$$

$$\forall \beta \sim_v \alpha \forall \gamma \sim_{\sigma \upharpoonright_n} \beta (\gamma \preceq \beta[\sigma \upharpoonright_n \mapsto \beta(v)] \longrightarrow S(\phi_{a-n+1}, \gamma \upharpoonright.)) \tag{4.5}$$

The equivalence between the first two is by Lemma 145. By compositional axioms the last sentence is equivalent to

$$\forall \beta \sim_v \alpha \forall \gamma \sim_{\sigma \upharpoonright_n} \beta \left( \gamma \preceq \beta[\sigma \upharpoonright_n \mapsto \beta(v)] \longrightarrow \forall \zeta \sim_{\sigma(n)} \gamma S(\phi_{a-n}, \zeta \upharpoonright.) \right)$$

Since  $\beta$  is universally quantified we can bind the size of  $\zeta$  using it and see that the above is in turn equivalent to

$$\forall \beta \sim_v \alpha \forall \gamma \sim_{\sigma \upharpoonright_n} \beta \forall \zeta \sim_{\sigma(n)} \gamma \left( \gamma \preceq \beta[\sigma \upharpoonright_n \mapsto \beta(v)] \wedge \zeta \preceq \gamma[\sigma(n) \mapsto \beta(v)] \longrightarrow S(\phi_{a-n}, \zeta \upharpoonright.) \right)$$

Now it can be easily seen that  $\gamma$  plays only an intermediary role and can be eliminated altogether from the above formula. Hence the above is equivalent to

$$\forall \beta \sim_v \alpha \forall \zeta \sim_{\sigma \upharpoonright_{n+1}} \beta (\zeta \preceq \beta[\sigma \upharpoonright_{n+1} \mapsto \beta(v)] \rightarrow S(\phi_{a-n}, \zeta \upharpoonright.))$$

Pushing quantifiers inside  $S$  (which we are allowed to do by Lemma 145), the above is equivalent to

$$\forall \beta \sim_v \alpha \ S(\text{bucl}(\sigma \upharpoonright_{n+1}, \phi_{a-n}, v), \beta \upharpoonright.)$$

Which by axiom for universal quantifier is equivalent to

$$S(\forall v \text{bucl}(\sigma \upharpoonright_{n+1}, \phi_{a-n}, v), \alpha)$$

which ends the proof. □

**Remark 148.** Let us observe that, in the proof, we did not make any important use of the general axiom for negation admissible in  $CS^-$ . Indeed, the only ingredients needed were the following statements:

1.  $\forall v \forall \phi(\bar{w}) \forall \alpha \in \text{Asn}(\ulcorner \forall v \phi \urcorner) \ (S(\ulcorner \forall v \phi \urcorner, \alpha) \equiv \forall \beta \sim_v \alpha S(\phi, \beta \upharpoonright_\phi))$
2.  $\forall v \forall \phi(\bar{w}) \forall \alpha \forall y \ \left( S(\forall v < \underline{y} \phi, \alpha) \equiv \forall \beta \sim_v \alpha (\beta(v) < y \rightarrow S(\phi, \beta \upharpoonright_\phi)) \right)$

and the presence of induction for  $\Delta_0$  formulae containing the truth predicate.

We obtain the following corollary:

**Corollary 149.** *The following sentence is provable already in  $PS_0$ :*

$$\forall\phi(\bar{w})\forall\sigma\forall\alpha \in \text{Asn}(\text{ucl}(\sigma, \phi)) \left( S(\text{ucl}(\sigma, \phi), \alpha) \equiv \forall\beta \sim_\sigma \alpha \ S(\phi, \beta \upharpoonright_\phi) \right)$$

**Corollary 150.** *Repeating the same proof for  $\exists$  instead of  $\forall$  one can show that  $PS_0$  (and  $CT_0$ . consequently) proves*

$$\forall\phi(\bar{w})\forall\sigma\forall\alpha \in \text{Asn}(\text{ecl}(\sigma, \phi)) \left( S(\text{ecl}(\sigma, \phi), \alpha) \equiv \exists\beta \sim_\sigma \alpha \ S(\phi, \beta \upharpoonright_\phi) \right)$$

Theorem 144 is really a key to using induction on the build-up of formulae almost freely. Proof of the theorem below neatly illustrates this technique.

**Theorem 151.** *The following sentence is provable in  $CS_0$ :*

$$\forall\phi(\bar{w})\forall\psi\forall\alpha \in \text{Asn}(\phi)\forall\beta \in \text{Asn}(\psi) (\phi[\alpha] = \psi[\beta] \rightarrow S(\phi, \alpha) \equiv S(\psi, \beta))$$

Let us note two important consequences of Theorem 151 first:

**Corollary 152.** *The following sentence is provable in  $CS_0$ :*

$$\forall\phi(\bar{w})\forall v\forall\alpha \in \text{Asn}(\forall v\phi) (S(\forall v\phi, \alpha) \equiv \forall x S(\phi[\underline{x}/v], \alpha))$$

*Proof.* Let us fix  $\phi, v$  and  $\alpha$  as in the assumptions. Then for every  $\beta \sim_v \alpha$ , such that  $\beta(v) = y$  we have

$$S(\phi, \beta \upharpoonright_\phi) \equiv S(\phi[\underline{y}/v], \alpha)$$

by Theorem 151 since  $\phi[\beta] = \phi[\underline{y}/v][\alpha]$ . Hence

$$\left( \forall\beta \sim_v \alpha \ S(\phi, \beta \upharpoonright_\phi) \right) \equiv \left( \forall y S(\phi[\underline{y}/x], \alpha) \right)$$

which is precisely what we wanted to prove.  $\square$

Before formulating next corollary, let us adopt one more convention: by saying, inside  $CS_0$ , that sentence  $\phi$  is *true*, we mean that it is satisfied by the empty assignment.

**Corollary 153.**  *$CS_0$  proves that all the axioms of PA are true. In particular, the following sentence is provable in  $CS_0$ :*

$$\forall\phi(\bar{w})\forall v \ (S(\text{Ind}(v, \phi), \emptyset))$$

where  $\text{Ind}(v, \phi)$  is as defined in Definition 29.

*Proof.* We work in  $CS_0$ . Let us fix  $\phi$  and a variable  $v$ . We shall start by eliminating the (possibly non-standard) quantifier prefix. Let us denote by  $\text{ind}(v, \phi)$  the formula  $\text{Ind}(v, \phi)$  with  $\forall\bar{y}$  omitted. By Theorem 144 we have:

$$S(\text{Ind}(v, \phi), \emptyset) \equiv \forall\beta \sim_{\bar{y}} \emptyset \ S(\text{ind}(v, \phi), \beta \upharpoonright_\phi)$$

Note that if  $\gamma$  in  $\text{Asn}(\forall v\phi)$  then,  $\gamma$  assigns values to  $\bar{y}$  with the exception of  $v$ . Hence, by Theorem 151 the right-hand side of the above is equivalent to

$$\forall \gamma \in \text{Asn}(\forall x\phi) \ S\left(\text{ind}(v, \phi[\gamma]), \emptyset\right).$$

Since  $\phi[\gamma]$  is a formula with precisely one free variable, without loss of generality we may consider only such formulae. This will reduce the complexity of expressions a little bit. So let  $\phi$  be any formula in which  $v$  is the only free variable. Assignments for  $\phi$  can be identified with numbers, hence we will freely use expressions such as

$$S(\phi, y)$$

meaning

$$\forall \beta \in \text{Asn}(\phi) \ \left( (\beta(v) = y) \rightarrow S(\phi, \beta) \right)$$

and so on. (Note that in the above  $x$  is fixed and *is not a free variable*. The only free variable is  $y$ .) Let us observe that the above is equivalent to a  $\Delta_0$  formula since the size of  $\beta$  can be bounded by  $\{\langle v, y \rangle\}$  (the function mapping  $x$  to  $y$ ). Let us observe that

$$S(\text{Ind}(\phi), \emptyset)$$

is, by compositional axioms and our conventions, equivalent to

$$\left[ \left( \forall y (S(\phi, y) \rightarrow S(\phi[v + 1/v], y)) \right) \rightarrow \left( S(\phi[\underline{0}/v], \emptyset) \rightarrow \forall y S(\phi, y) \right) \right]$$

Moreover, by Theorem 151

$$\forall y \ \left( S(\phi[v + 1/v], y) \equiv S(\phi, y + 1) \right)$$

and

$$S(\phi[\underline{0}/v], \emptyset) \equiv S(\phi, 0)$$

Hence the above is equivalent to

$$\left( \forall y (S(\phi, y) \rightarrow S(\phi, y + 1)) \right) \rightarrow \left( S(\phi, 0) \rightarrow \forall y S(\phi, y) \right)$$

which is an axiom of  $\text{CS}_0$ , since  $S(\phi, y)$  is a  $\Delta_0$  formula. □

**Remark 154.** Equivalently, one could reduce  $S(\text{Ind}(\phi(x)), \emptyset)$  to

$$\forall y \ \left( S(\phi(\underline{y}), \emptyset) \rightarrow S(\phi(\underline{y} + 1), \emptyset) \right) \rightarrow \left( S(\phi(\underline{0}), \emptyset) \rightarrow \forall y \ S(\phi(\underline{y}), \emptyset) \right)$$

which is more natural if one thinks of  $S(\phi, \emptyset)$  as of a truth predicate. We chose the other strategy because we think of  $S(\phi, x)$  as of a satisfaction predicate, but we treat it as a purely aesthetical choice.

We shall now prove Theorem 151. Before that we need some preparations:

**Definition 155 (PA).** A *generalised numeral* is any term of the form

$$1 + (1 + (1 + \dots (1 + x)))$$

where  $x$  is a variable or 0. In particular, each variable is a generalised numeral. In other words, a generalised numeral is any term that

1. contains exactly one free variable
2. if we substitute an arbitrary *numeral* for this free variable, then the result is a *numeral*.

If  $s$  is a generalised numeral, then the *length* of  $s$  is the number of 0 and 1 symbol used in  $s$ .

Let  $\phi$  be a formula. An occurrence of a term  $l_\pi^\phi$  is called a *bounded generalised numeral* if  $l^\phi(\pi)$  is a generalised numeral, but not a numeral and some prefix of  $\pi$  is labelled by  $l^\phi$  with a formula starting with  $\exists v$ . Otherwise the occurrence of a generalised numeral in  $\phi$  is called *free*.

**Example 156.**  $1 + (1 + (1 + 0))$  has length 4, and  $1 + (1 + (1 + v))$  has length 3. The only generalised numerals of length 0 are variables. Any occurrence of a closed term in a formula is a free generalised numeral.

**Proposition 157.** If  $\phi, \psi$  are two formulae such that for some  $\alpha, \beta$ ,

$$\phi[\alpha] = \psi[\beta]$$

then there exists third formula  $\omega$  such that

1. for some assignment  $\gamma$ ,  $\omega[\gamma] = \phi[\alpha] = \psi[\beta]$ ;
2.  $FV(\omega) \cap FV(\phi) = \emptyset$ ,  $FV(\omega) \cap FV(\psi) = \emptyset$ ;
3. the only free generalised numerals occurring in  $\omega$  are variables.

*Proof.* We work in PA. Let  $v$  be any variable that does not occur in  $\phi$ . We will treat  $v$  as an additional free variable for marking places for *numerals* in a formula. Formally: let  $\phi'$  result from  $\phi$  by formally substituting  $v$  for every free variable of  $\phi$ . Now, let  $\bar{\phi}$  result from  $\phi'$  by formally substituting  $v$  for every (occurrence of a) term which is either a numeral or a free generalised numeral which contains  $v$  as the only free variable. Define  $\omega$  to be the formula resulting from substituting  $v_{i_1}, v_{i_2}, v_{i_3}, \dots$  in  $\bar{\phi}$  for the first, the second, the third,  $\dots$  occurrence of  $v$ , respectively, where the respective ordering of occurrences of free variables is  $\prec_\phi$  defined in Definition 21 and  $v_{i_1}, v_{i_2}, v_{i_3}, \dots$  are variables with consecutive least possible indices that do not occur (either as free or bounded ones) in  $\phi$  or  $\psi$ . So defined  $\omega$  satisfies 2 and 3 in the thesis of our proposition. Additionally it has the following property: each variable that occurs as a free variable in  $\omega$ , occurs exactly once in  $\omega$ . We define  $\gamma$ : for  $j > i$  let  $l_{\pi_j}^\omega$  denote the occurrence of a variable  $v_j$  in  $\omega$ . Let us put

$$\gamma(v_j) = \left( l_{\pi_j}^\omega \right)_\alpha^\circ$$

Where  $l^\phi$  is as defined in Definition 20. Let us observe that  $\omega$  and  $\gamma$  are uniquely defined in terms of  $\phi, \psi$  and  $\alpha$ . Let us denote the respective functions by  $\omega(\phi, \psi)$  and  $\gamma(\phi, \psi, \alpha)$ . Let us

check that  $\omega[\gamma] = \phi[\alpha]$ . Let  $n_1^\phi$  denote the number of free generalised numerals in  $\phi$  of maximal length and  $n_2^\phi$  the maximal length of a free generalised numeral occurring in  $\phi$ . The proof proceeds of by induction on

$$n^\phi = \max\{n_1^\phi, n_2^\phi\}$$

If  $n^\phi = 0$ , then the only free numerals in  $\phi$  are free variables. Consequently, for any  $\psi, \alpha, \beta$  such that  $\phi[\alpha] = \psi[\beta]$ ,  $\omega(\phi, \psi)$  differs from  $\phi$  at most with respect to shapes of free variables. In particular, the full skeletons (as defined in Definition 20) of  $\omega(\phi, \psi)$  and  $\phi$  are the same. Let us abbreviate  $\omega(\phi, \psi)$  with  $\omega$  and  $\gamma(\phi, \psi, \alpha)$  with  $\gamma$ . To see that  $\omega[\gamma] = \phi[\alpha]$  it is enough to check if  $l_\pi^\omega$  is a free occurrence of a variable, then  $\gamma(l^\omega(\pi)) = \alpha(l^\phi(\pi))$ . But this is clear from the definition of  $\gamma$ .

Let us suppose that  $n^\phi = k + 1$  and for any  $\phi, \psi, \alpha$ , if  $n^\phi < k + 1$  then  $\omega(\phi, \psi)[\gamma(\phi, \psi, \alpha)] = \phi[\alpha]$ . Let  $l_\pi^\phi$  be the  $\prec_\phi$ -least occurrence of a maximal free numeral in  $\phi$ . If  $l^\phi(\pi)$  contains a variable which occurs in  $\phi$  exactly once, then let  $v$  be that variable. Otherwise, let  $v$  be a variable which does not occur in  $\phi, \omega$  and  $\psi$ . Define  $\phi'$  to be the result of substituting  $v$  for  $l_\pi^\phi$ . Now  $\omega(\phi', \psi) = \omega(\phi, \psi)$  and  $n^{\phi'} < k + 1$ . Define the assignment  $\alpha'$

$$\begin{aligned} \text{dom}(\alpha') &= \text{dom}(\alpha) \cup \{v\} \\ \alpha'(x) &= \alpha(x) \text{ if } x \neq v \\ \alpha'(v) &= \left(l_\pi^\phi\right)_\alpha^\circ \end{aligned}$$

Now we have  $\phi'[\alpha'] = \phi[\alpha]$  and  $\gamma(\phi', \psi, \alpha') = \gamma(\phi, \psi, \alpha)$ . The thesis follows by our induction assumption.  $\square$

Before stating the next proposition, let us have a word of comment which will clarify its meaning: given a binary sequence  $\pi$  and formulae  $\phi, \psi$  let us say that  $\pi$  determines  $\psi$  in  $\phi$  if and only if  $l^\phi(\pi) = \psi$ . Then the next proposition says, *inter alia*, that if  $\pi$  determines a disjunction in  $\omega(\phi, \psi)$ , then it determines a disjunction in  $\phi$ , as well.

**Proposition 158 (PA).** *Let  $\phi, \psi$  be two formulae and  $\omega = \omega(\phi, \psi)$  be the formula from Proposition 157. Then the following hold*

1. if  $l_\pi^\omega$  is an occurrence of bounded variable in  $\omega$ , then  $l^\phi(\pi) = l^\omega(\pi)$
2. if  $l_\pi^\omega$  is an occurrence of free variable in  $\omega$ , then  $l_\pi^\phi$  is an occurrence of a free numeral in  $\phi$ .
3. if  $\odot \in \{\cdot, +\}$  and  $l^\omega(\pi) = l^\omega(\pi \frown 0) \odot l^\omega(\pi \frown 1)$  then  $l^\phi(\pi) = l^\phi(\pi \frown 0) \odot l^\phi(\pi \frown 1)$ ;
4. if  $l_\pi^\omega$  is any occurrence of a formula of the form  $s = t$  in  $\omega$ , then

$$l^\phi(\pi) = R(l^\phi(\pi \frown 0), l^\phi(\pi \frown 1))$$

5. if  $l_\pi^\omega$  is any occurrence of a formula of the form  $\theta_0 \vee \theta_1$  in  $\omega$ , then

$$l^\phi(\pi) = l^\phi(\pi \frown 0) \vee l^\phi(\pi \frown 1)$$

6. if  $l_\pi^\omega$  is any occurrence of a formula of the form  $\neg\theta$  in  $\omega$ , then

$$l^\phi(\pi) = \neg l^\phi(\pi \frown 0)$$

7. if  $l_\pi^\omega$  is any occurrence of a formula of the form  $\exists v\theta$  in  $\omega$ , then

$$l^\phi(\pi) = \exists v l^\phi(\pi \frown 0)$$

**Definition 159 (PA).** Let  $\phi$  be a formula and  $l_\pi^\phi$  an occurrence of its subformula or a term.  $\sigma^\phi(\pi)$  is a sequence of variables (listed in the order of increasing indices) such that  $v$  appears in  $\sigma^\phi(\pi)$  if and only if for some  $\rho$

1.  $l_\rho^{l^\phi(\pi)}$  is a free occurrence of  $v$  in  $l^\phi(\pi)$  and
2.  $l_{\pi \frown \rho}^\phi$  is a bounded occurrence of  $v$  in  $\phi$ .

**Example 160.** Let  $\phi = \exists v_1 \exists v_2 (v_1 = 0 \wedge \exists v_0 (v_0 = v_1 + v_2)) \vee \exists v_0 (v_0 = v_1 + v_2)$ . Let

$$\begin{aligned} \pi &= [0, 0, 0, 1] \\ \pi' &= [1] \\ \rho &= [0, 0, 0, 1, 0, 1, 1] \\ \rho' &= [1, 0, 1, 1] \end{aligned}$$

Then  $l^\phi(\pi) = \exists v_0 (v_0 = v_1 + v_2) = l^\phi(\pi')$ , and  $l^\phi(\rho) = v_2 = l^\phi(\rho')$  but

$$\begin{aligned} \sigma^\phi(\pi) &= [v_1, v_2] \\ \sigma^\phi(\pi') &= \varepsilon \\ \sigma^\phi(\rho) &= [v_2] \\ \sigma^\phi(\rho') &= \varepsilon \end{aligned}$$

The proposition below is simply a generalisation of point 1 from Proposition 158:

**Proposition 161.** Let  $\phi, \psi$  be arbitrary formulae and let  $\omega = \omega(\phi, \psi)$ . Let  $l_\pi^\omega$  be an arbitrary occurrence of a subformula in  $\omega$ . Then  $\sigma^\phi(\pi) = \sigma^\omega(\pi)$ .

*Proof of Theorem 151.* We work in  $\text{CS}_0$ . Fix  $\phi$  and  $\psi$  and  $\alpha \in \text{Asn}(\phi), \beta \in \text{Asn}(\psi)$  such that

$$\phi[\alpha] = \psi[\beta]$$

Let  $\omega$  and  $\gamma$  be a formula and an assignment from Proposition 157 We shall show that  $S(\omega, \gamma) \equiv S(\phi, \alpha)$  which, by symmetricity of roles played by pairs  $(\phi, \alpha), (\psi, \beta)$ , will suffice to end the proof.

Since  $\text{FV}(\phi) \cap \text{FV}(\omega) = \emptyset$ , then

$$\text{dom}(\alpha) \cap \text{dom}(\gamma) = \emptyset$$

Hence by compositional axioms of  $\text{CS}^-$  we have

$$(S(\omega, \gamma) \equiv S(\phi, \alpha)) \equiv S(\omega \equiv \phi, \gamma \cup \alpha)$$

where  $\gamma \cup \alpha$  is a set-theoretic sum of two functions (by assumption no clashes are possible). For our convenience, let us denote  $\gamma \cup \alpha$  by  $\delta$ . We have

$$\omega[\delta] = \phi[\delta]$$

Let  $\sigma^\phi$  and  $\sigma^\omega$  be as defined in Definition 159. By Proposition 161 we may skip the superscript above  $\sigma$ . Let  $k$  be the complexity of  $\omega$ . By induction on  $n$  in the formula

$$\forall \pi \in \text{Skel}(\omega) \left( \text{len}(\pi) \geq k - n \rightarrow S\left(\text{ucl}(\sigma(\pi), l^\omega(\pi) \equiv l^\phi(\pi)), \delta \uparrow \right) \right) \quad (4.6)$$

we show that

$$\forall \pi \in \text{Skel}(\omega) \ S\left(\text{ucl}(\sigma(\pi), l^\omega(\pi) \equiv l^\phi(\pi)), \delta \uparrow \right)$$

Showing this our proof will be finished since:

1.  $\varepsilon$  is the unique sequence in  $\text{Skel}(\omega)$  of length 0
2.  $\sigma(l(\varepsilon))$  is the empty sequence and
3.  $l^\phi(\varepsilon) = \phi, l^\omega(\varepsilon) = \omega$ .

$\Delta_0$  induction suffices for this purpose, since, by collection we know that there exists  $u$  such that

$$\forall \pi < \text{Skel}(\omega) \ \langle \text{ucl}(\sigma(\pi), l^\omega(\pi) \equiv l^\phi(\pi)), \delta \rangle < u$$

Hence we can apply Remark 57 to justify the use of induction for 4.6.ss In the base step for  $n = 0$  we have to show that if  $\pi$  is of length  $k$ , then

$$S\left(\text{ucl}(\sigma(\pi), l^\omega(\pi) \equiv l^\phi(\pi)), \delta \uparrow \right).$$

If  $\text{len}(\pi) = k$ , then  $l^\omega_\pi$  and  $l^\phi_\pi$  are atomic formulae. Suppose  $l^\omega(\pi) = (s = t)$  for some terms  $s, t$ . Then  $l^\phi(\pi) = (l^\phi(\pi \frown 0) = l^\phi(\pi \frown 1))$ . Observe that  $s, t$  and hence  $l^\phi(\pi \frown 0), l^\phi(\pi \frown 1)$  might contain a non-standard number of variables which are bounded by a quantifier in  $\omega$  (and hence in  $\phi$ ). We have to show that for every  $\beta \sim_{\lambda(\pi)} \delta, \beta \in \text{Asn}((s = t) \equiv (l^\phi(\pi \frown 0) = l^\phi(\pi \frown 1)))$

$$\left( (s)_\beta^\circ = (t)_\beta^\circ \right) \equiv \left( (l^\phi(\pi \frown 0))_\beta^\circ = (l^\phi(\pi \frown 1))_\beta^\circ \right)$$

This can be shown by induction on the build-up of  $s$  and  $t$  (the inductive step is trivial and the base step uses the assumptions on  $\beta$  and  $\omega$ ).

Let us fix  $0 < n < k$  and assume the thesis holds for  $n$ . Let us fix  $\pi$  of length  $k - (n + 1)$ . We shall proceed in two steps, since the step for  $\neg$  is trivial.

*Step 1:* Let  $l^\omega(\pi) = l^\omega(\pi \frown 0) \vee l^\omega(\pi \frown 1)$ , then  $l^\phi(\pi) = l^\phi(\pi \frown 0) \vee l^\phi(\pi \frown 1)$ . To enhance readability let us denote  $\pi \frown i$  by  $\pi_i$  ( $i = 0, 1$ ) and abbreviate

$$\begin{aligned} l^\omega(\pi) \text{ with } \theta^\omega \text{ and } l^\phi(\pi) \text{ with } \theta^\phi \\ l^\omega(\pi_i) \text{ with } \theta_i^\omega \text{ and } l^\phi(\pi_i) \text{ with } \theta_i^\phi \end{aligned}$$

By induction assumption ( $\text{len}(\pi_0) = \text{len}(\pi_1) = k - (n + 1) + 1 = k - n$ ).

$$S\left(\text{ucl}(\sigma(\pi_i), \theta_i^\omega \equiv \theta_i^\phi), \delta \upharpoonright \cdot\right)$$

where  $i = 0, 1$ . The above is equivalent to

$$\forall \zeta \sim_{\sigma(\pi_i)} \delta \upharpoonright_{\theta_i^\omega \equiv \theta_i^\phi} \left( S(\theta_i^\omega, \zeta \upharpoonright \cdot) \equiv S(\theta_i^\phi, \zeta \upharpoonright \cdot) \right)$$

We have to show that the above implies

$$\forall \zeta \sim_{\sigma(\pi)} \delta \upharpoonright_\theta \left( (S(\theta_0^\omega, \zeta \upharpoonright \cdot) \vee S(\theta_1^\omega, \zeta \upharpoonright \cdot)) \equiv (S(\theta_0^\phi, \zeta \upharpoonright \cdot) \vee S(\theta_1^\phi, \zeta \upharpoonright \cdot)) \right)$$

which is clearly the case. Observe that e.g. if  $\zeta \sim_{\sigma(\pi)} \delta \upharpoonright_{\theta^\omega \equiv \theta^\phi}$ , then

$$\zeta \upharpoonright_{\theta_i^\omega \equiv \theta_i^\phi} \sim_{\sigma(\pi_i)} \delta \upharpoonright_{\theta_i^\omega \equiv \theta_i^\phi}.$$

*Step 2:* Now consider the case of  $l^\omega(\pi) = \exists v l^\omega(\pi \frown 0)$ . As in the above abbreviate  $\pi \frown 0$  with  $\pi_0$  and

$$\begin{aligned} & l^\omega(\pi) \text{ with } \theta^\omega \text{ and } l^\phi(\pi) \text{ with } \theta^\phi \\ & l^\omega(\pi_0) \text{ with } \theta_0^\omega \text{ and } l^\phi(\pi_0) \text{ with } \theta_0^\phi \end{aligned}$$

Then  $\theta^\phi = \exists v \theta_0^\phi$ , and, consequently,  $v$  is either listed in sequence  $\sigma(\pi_0)$  or is neither a free variable of  $\theta_0^\omega$  nor of  $\theta_0^\phi$ . In both cases we have

$$\text{im}((\sigma(\pi))) = \text{im}(\sigma(\pi_0)) \setminus \{v\}.$$

By inductive assumption we have (pulling the quantifier prefix outside)

$$\forall \zeta \sim_{\sigma(\pi_0)} \delta \upharpoonright_{\theta_0^\omega \equiv \theta_0^\phi} \left( S(\theta_0^\omega, \zeta \upharpoonright \cdot) \equiv S(\theta_0^\phi, \zeta \upharpoonright \cdot) \right)$$

Every  $\zeta \sim_{\sigma(\pi_0)} \delta \upharpoonright_{\theta_0^\omega \equiv \theta_0^\phi}$  can be decomposed into  $\zeta_1 \sim_{\sigma(\pi_0)-v} \delta \upharpoonright_{\theta^\omega \equiv \theta^\phi}$  and  $\zeta_2 \sim_v \zeta_1$ , so we have

$$\forall \zeta_1 \sim_{\sigma(\pi_0)-v} \delta \upharpoonright_{\theta^\omega \equiv \theta^\phi} \forall \zeta_2 \sim_v \zeta_1 \left( S(\theta_0^\omega, \zeta_2 \upharpoonright \cdot) \equiv S(\theta_0^\phi, \zeta_2 \upharpoonright \cdot) \right)$$

The above implies

$$\forall \zeta_1 \sim_{\sigma(\pi)} \delta \upharpoonright_{\theta^\omega \equiv \theta^\phi} \left( \exists \zeta_2 \sim_v \zeta_1 \ S(l^\omega(\pi_0), \zeta_2 \upharpoonright \cdot) \equiv \exists \zeta_2 \sim_v \zeta_1 \ S(l^\phi(\pi_1), \zeta_2 \upharpoonright \cdot) \right)$$

which by the compositional axiom for existential quantifier is equivalent to

$$\forall \zeta_1 \sim_{\sigma(\pi)} \delta \upharpoonright_{\theta^\omega \equiv \theta^\phi} \left( S(\theta^\omega, \zeta_1 \upharpoonright \cdot) \equiv S(\theta^\phi, \zeta_1 \upharpoonright \cdot) \right)$$

This ends the whole proof. □

**Remark 162.** The crucial idea used in the above proof is that by Theorem 144 we can reduce  $\Pi_1$ -inductive assumption of type

$$\forall \beta \sim_\sigma \delta \ S(\phi, \beta)$$

to the  $\Delta_0$ -formula

$$S(\text{ucl}(\sigma, \phi), \delta),$$

where  $\text{ucl}(\sigma, \phi)$  may be taken as a parameter.

The next regularity property solves the problem of term regularity. Let us adopt the following conventions:

**Definition 163 (PA).** Let  $\phi$  be a formula and let  $\sigma$  be any injective enumeration of some variables that have free occurrences in  $\phi$ . Let  $\beta$  be any sequence of closed terms of the same length as  $\sigma$  (the set of such sequences will be denoted by  $\text{Terms}^\sigma$ . Formally,  $\sigma$  is a parameter in this formula). Then

1. By  $\beta^\circ$  we mean a sequence of values of terms belonging to  $\beta$ .
2. By  $\phi[\sigma/\beta]$  we mean a formula resulting from substituting  $\beta(i)$  for  $\sigma(i)$  in  $\phi$ .

**Theorem 164 (Term Regularity).** *The following sentence is provable in  $\text{CS}_0$*

$$\forall \phi(\bar{w}) \forall \tau \forall \alpha, \beta \left( \text{im}(\tau) \subseteq \text{FV}(\phi) \wedge \alpha \in \text{Terms}^\tau \wedge \beta \in \text{Terms}^\tau \wedge \alpha = \beta^\circ \longrightarrow \right. \\ \left. \longrightarrow \forall \gamma \in \text{Asn}(\phi) \left( S(\phi[\tau/\alpha], \gamma \upharpoonright_{\phi[\tau/\alpha]}) \equiv S(\phi[\tau/\beta], \gamma \upharpoonright_{\phi[\tau/\beta]}) \right) \right) \quad (4.7)$$

*Proof.* We apply the same strategy as in the proof of Theorem 151. Let  $\phi, \tau, \alpha, \beta$  satisfy the antecedent of the above implication. Let us fix  $\gamma$ . Let us observe that

$$\text{Skel}(\phi) = \text{Skel}(\phi[\tau/\alpha]) = \text{Skel}(\phi[\tau/\beta])$$

For  $\pi \in \text{Skel}(\phi)$  let  $\sigma^\phi(\pi)$  be the sequence from Definition 159. Let us observe that

$$\sigma^{\phi[\tau/\alpha]}(\pi) \equiv \sigma^{\phi[\tau/\beta]}(\pi)$$

for every  $\pi \in \text{Skel}(\phi)$ , so we shall write simply  $\sigma(\pi)$  to denote any of the above. Let  $k$  be the complexity of  $\phi$ . By induction on  $n$  in the formula

$$\forall \pi \in \text{Skel}(\phi) \left( \text{len}(\pi) \geq k - n \rightarrow S(\text{ucl}(\sigma(\pi), l^{\phi[\tau/\alpha]}(\pi) \equiv l^{\phi[\tau/\beta]}(\pi)), \gamma \upharpoonright_{\cdot}) \right)$$

we show

$$\forall \pi \in \text{Skel}(\phi) \left( S(\text{ucl}(\sigma(\pi), l^{\phi[\tau/\alpha]}(\pi) \equiv l^{\phi[\tau/\beta]}(\pi)), \gamma \upharpoonright_{\cdot}) \right)$$

imitating the proof of Theorem 151. □

The last regularity property we would like to prove is the closure under  $\alpha$ -conversion. Let us recall that  $\phi$  and  $\psi$  are  $\alpha$  equivalent (denoted  $\phi \equiv_\alpha \psi$ ) if and only if  $\phi$  can be obtained from  $\psi$  by, and only by, renumbering bounded variables (but no free occurrence of a variable may become bounded). The following lemma is a variant of Proposition 158:

**Proposition 165 (PA).** Let  $\phi, \psi$  be two formulae such that  $\phi \equiv_{\alpha} \psi$ . Let  $\chi$  be the relevant permutation of variables. Then the following hold

1. if  $l_{\pi}^{\phi}$  is an occurrence of bounded variable in  $\phi$ , then  $l^{\psi}(\pi) = \chi(l^{\phi}(\pi))$ ;
2. if  $l_{\pi}^{\phi}$  is an occurrence of a free numeral in  $\phi$ , then  $l^{\psi}(\pi) = l^{\phi}(\pi)$ ;
3. if  $\odot \in \{\cdot, +\}$  and  $l^{\phi}(\pi) = l^{\phi}(\pi \frown 0) \odot l^{\phi}(\pi \frown 1)$  then  $l^{\psi}(\pi) = l^{\psi}(\pi \frown 0) \odot l^{\psi}(\pi \frown 1)$ ;
4. if  $l_{\pi}^{\phi}$  is any occurrence of a formula of the form  $s = t$  in  $\phi$ , then

$$l^{\psi}(\pi) = (l^{\psi}(\pi \frown 0) = l^{\psi}(\pi \frown 1));$$

5. if  $l_{\pi}^{\phi}$  is any occurrence of a formula of the form  $\theta_0 \vee \theta_1$  in  $\phi$ , then

$$l^{\psi}(\pi) = l^{\psi}(\pi \frown 0) \vee l^{\psi}(\pi \frown 1);$$

6. if  $l_{\pi}^{\phi}$  is any occurrence of a formula of the form  $\neg\theta$  in  $\phi$ , then

$$l^{\psi}(\pi) = \neg l^{\psi}(\pi \frown 0);$$

7. if  $l_{\pi}^{\phi}$  is any occurrence of a formula of the form  $\exists v\theta$  in  $\phi$ , then

$$l^{\psi}(\pi) = \exists \chi(v) l^{\psi}(\pi \frown 0).$$

In the proof of the  $\alpha$ -correctness theorem we will need one more, easily provable, trait of  $\text{CS}_0$ : it is stated in the next proposition:

**Proposition 166.**  $\text{CS}_0 \vdash \text{DC}(S)$

*Proof.* We work in  $\text{CS}_0$ . Let us fix a set of sentences  $c$  and an assignment  $\alpha \in \text{Asn}(c)$ . Let  $y = c \upharpoonright_n$  denote the arithmetical formula representing the relation: " $y$  is a set of first  $n$  elements of  $c$ ". We use  $\Delta_0$  induction on  $y$  in the formula

$$\theta(y) := \forall n < y \forall d < c \left( d = c \upharpoonright_n \rightarrow S \left( \bigvee_{\phi \in d} \phi, \alpha \upharpoonright_{\cdot} \right) \equiv (\exists \phi \in d S(\phi, \alpha \upharpoonright_{\cdot})) \right)$$

The proof of the base step is trivial and the proof of the induction step uses a routine argument, for if  $d$  consists of first  $y$  elements of  $c$ , then, by Definition 131  $\bigvee_{\phi \in d} \phi$  is equal to  $\psi \vee \bigvee_{\phi \in e} \phi$ , where  $\psi = \max(d)$  and  $e = d \setminus \{\psi\}$ . Now if  $\psi$  is not satisfied by  $\alpha \upharpoonright_{\psi}$ , then  $\bigvee_{\phi \in e} \phi$  must be satisfied by  $\alpha \upharpoonright_{\bigvee_{\phi \in e} \phi}$  and we may use our induction assumption, since  $e = c \upharpoonright_{y-1}$ .  $\square$

**Corollary 167.** The following sentence is provable in  $\text{CS}_0$

$$\forall c \left( \text{SetSent}(c) \rightarrow \forall \alpha \in \text{Asn}(c) \left( S \left( \bigwedge_{\phi \in c} \phi, \alpha \right) \equiv \forall \phi \in c S(\phi, \alpha \upharpoonright_{\phi}) \right) \right)$$

Now we can start proving the theorem on  $\alpha$ -correctness.

**Theorem 168.** *The following sentence is provable in  $CS_0$*

$$\forall \phi(\bar{w}), \psi(\bar{w}) \forall \alpha \in \text{Asn}(\phi) \left( \phi \equiv_{\alpha} \psi \rightarrow S(\phi, \alpha) \equiv S(\psi, \alpha) \right) \quad (4.8)$$

*Proof.* Working in  $CS_0$  let us fix  $\phi, \psi$  and  $\alpha$  as required. We shall start with two reductions. Firstly, by Theorem 151 it is sufficient to prove our theorem for  $\phi, \psi$  such that the set of variables which have free occurrences in  $\phi$  (equivalently in  $\psi$ ) is disjoint from the set of variables which have bounded occurrences either in  $\phi$  or in  $\psi$ . This is because we can exchange free variables in  $\phi$  and  $\psi$  changing  $\alpha$  analogously. Secondly, we can demand that the set of variables which have bounded occurrences in  $\phi$  is disjoint from the set of variables which have bounded occurrences in  $\psi$ . This is because for every  $\phi, \psi$  we can always find third formula  $\theta$  such that pairs

$$(\phi, \theta), (\psi, \theta)$$

satisfy this requirement and argue similarly as in the proof of Theorem 151. So let us assume that  $\phi$  and  $\psi$  satisfy both requirements. Let  $\chi$  be a bijection between the set of bounded variables of  $\phi$  and  $\psi$  (by our assumptions on variables of  $\phi$  and  $\psi$  we can stop talking about occurrences in such contexts). For  $\pi \in \text{Skel}(\phi) = \text{Skel}(\psi)$  let  $\sigma^{\phi}(\pi)$  be a sequence defined in Definition 159. Then

$$\chi \circ (\sigma^{\phi}(\pi)) = \sigma^{\psi}(\pi).$$

Let  $k$  be the complexity of  $\phi$ . Let  $\sigma(\pi)$  abbreviate  $\sigma^{\phi}(\pi) \frown \sigma^{\psi}(\pi)$  and  $\theta(\pi)$  abbreviate the formula:

$$\text{ucl}(\sigma(\pi), \left( \bigwedge_{v_i \in \sigma^{\phi}(\pi)} v_i = \chi(v_i) \right) \rightarrow l^{\phi}(\pi) \equiv l^{\psi}(\pi))$$

Let us observe that for every  $\alpha \in \text{Asn}(\theta(\pi))$   $S(\theta(\pi), \alpha \upharpoonright_{\cdot})$  is equivalent to

$$\forall \beta \sim_{\sigma(\pi)} \alpha \upharpoonright_{\theta(\pi)} \left( \forall v_i \in \sigma^{\phi}(\pi) (\beta(v_i) = \beta(\chi(v_i))) \rightarrow (S(l^{\phi}(\pi), \beta \upharpoonright_{\cdot}) \equiv S(l^{\psi}(\pi), \beta \upharpoonright_{\cdot})) \right)$$

By induction on  $n$  in the formula

$$\gamma(n) := \forall \pi \in \text{Skel}(\phi) \left( \text{len}(\pi) \geq k - n \rightarrow S(\theta(\pi), \alpha \upharpoonright_{\cdot}) \right)$$

we show that

$$\forall \pi \in \text{Skel}(\phi) \left( S(\theta(\pi), \alpha \upharpoonright_{\cdot}) \right)$$

The argument is fully analogous to that given in Theorem 151. In the base step for  $n = 0$ , when  $l^{\phi}(\pi)$  and  $l^{\psi}(\pi)$  have to be atomic, we apply subinduction on the complexity of terms. More precisely, if  $A^{\phi}$  is the full tree of  $\phi$  (as in Definition 20) and  $m$  the maximal length of its path then by induction on  $n$  in

$$\begin{aligned} \forall \pi \in A^{\phi} \left( l^{\phi}(\pi) \in \text{Terms}(\phi) \wedge \text{len}(\pi) \geq m - n \rightarrow \right. \\ \left. \rightarrow S \left( \text{ucl}(\sigma(\pi), \left( \bigwedge_{v_i \in \sigma^{\phi}(\pi)} v_i = \chi(v_i) \right) \rightarrow l^{\phi}(\pi) = l^{\psi}(\pi)), \alpha \upharpoonright_{\cdot} \right) \right) \end{aligned}$$

we show

$$\forall \pi \in A^\phi \left( l^\phi(\pi) \in \text{Terms}(\phi) \longrightarrow S \left( \text{ucl}(\sigma(\pi), \left( \bigwedge_{v_i \in \sigma^\phi(\pi)} v_i = \chi(v_i) \right) \rightarrow l^\phi(\pi) = l^\psi(\pi)), \alpha \uparrow. \right) \right)$$

Let us show one step in the proof that  $\forall n(\gamma(n) \rightarrow \gamma(n+1))$ . Let us fix  $\pi$  of length  $k - (n+1)$  and assume that  $l^\phi(\pi) = \exists v l^\phi(\pi \frown 0)$ . If  $v \notin \sigma^\phi(\pi \frown 0)$ , then this case can be easily covered by our induction assumption. Assume that  $v \in \sigma^\phi(\pi \frown 0)$ . In such a case  $v \notin \sigma^\phi(\pi)$ . Let us abbreviate  $\pi \frown 0$  with  $\pi_0$ . We have to demonstrate that

$$\begin{aligned} \forall \beta \sim_{\sigma(\pi)} \alpha \uparrow_{\theta(\pi)} \left( \forall v_i \in \sigma^\phi(\pi) (\beta(v_i) = \beta(\chi(v_i))) \longrightarrow \right. \\ \left. \longrightarrow \left( (\exists \gamma \sim_v \beta \uparrow_{l^\phi(\pi_0)} S(l^\phi(\pi_0), \gamma \uparrow.)) \equiv (\exists \gamma \sim_{\chi(v)} \beta \uparrow_{l^\psi(\pi_0)} S(l^\psi(\pi_0), \gamma \uparrow.)) \right) \right) \end{aligned}$$

Let us fix  $\beta$  satisfying the antecedent of the above implication. Assume that

$$\exists \gamma \sim_v \beta \uparrow_{l^\psi(\pi_0)} S(l^\psi(\pi_0), \gamma \uparrow.)$$

Since  $v \in \text{BV}(\phi)$ , then by our assumption on  $\phi$ ,  $v \notin \text{FV}(\phi)$  and therefore  $v \notin \text{dom}(\beta)$ . Let us fix  $\gamma$  witnessing the above existential quantifier. Since  $\chi(v) \in \text{BV}(\psi)$ , then  $\chi(v) \notin \text{dom}(\gamma)$ . Let us define assignment  $\gamma'$  with domain  $\text{dom}(\gamma) \cup \{\chi(v)\}$  by putting

$$\gamma'(w) = \begin{cases} \gamma(w) & \text{if } w \neq \chi(v) \\ \gamma(v) & \text{if } w = \chi(v) \end{cases}$$

Then we have

$$\gamma' \sim_{\sigma(\pi_0)} \beta$$

and

$$\forall v_i \in \sigma^\phi(\pi_0) (\gamma'(v_i) = \gamma'(\chi(v_i)))$$

So we may use our induction assumption ( $\text{len}(\pi_0) = k - n$ ) and conclude that

$$S(l^\psi(\pi_0), \gamma' \uparrow.)$$

Let us observe that  $\gamma' \uparrow_{l^\psi(\pi_0)} \sim_{\chi(v)} \beta$ . Indeed  $\gamma' \sim_{\chi(v)} \gamma \sim_v \beta$ , hence

$$\gamma' \sim_{\chi(v), v} \beta$$

Hence  $\gamma' \uparrow_{l^\psi(\pi_0)} \sim_{\chi(v)} \beta \uparrow_{l^\psi(\pi_0)}$  and our proof is finished.  $\square$

We are ready to prove Theorem 142:

*Proof of Theorem 142.* We have already shown that, provably in  $\text{CS}_0$ , every axiom of PA is true. What is left to show is that every consequence of those axioms is true; i.e. the following sentence:

$$\forall \phi \text{ Pr}_{\text{PA}}(\phi) \rightarrow \forall \alpha \in \text{Asn}(\phi) S(\phi, \alpha)$$

We have to show that generalisations of all formulae of the form 1-6 from Definition 41 are, provably in  $CS_0$ , satisfied by every assignment for them. By Theorem 144 it is sufficient that every formula from 1-6 is satisfied by every assignment for it. For 1 this is clear from propositional axioms (if we want only axioms for propositional calculus) or the fact that by  $\Delta_0$  induction we can perform the induction on the length of proofs. Theorem 151 and Theorem 164 together take care of 2 and 6. Compositional axioms themselves are enough to assure 3-5. Now, let  $\phi$  be a PA-provable formula and let  $\alpha \in \text{Asn}(\phi)$ . Let  $d = \phi_0, \dots, \phi_a = \phi$  be a proof of  $\phi$ . Let  $\sigma(\phi_i)$  be the sequence of variables which have free occurrence in  $\phi_i$  and are not in the domain of  $\alpha$  (by assumption  $\sigma(\phi) = \emptyset$ ). By induction on  $n$  up to  $a$ , we show that

$$\forall i \leq a \ S(\text{ucl}(\sigma(\phi_i), \phi_i), \alpha \upharpoonright_i)$$

we use Remark 57 to justify that the induction axiom for the above formula is provable in  $CS_0$ . The base step was verified above. To do the induction step assume

$$S(\text{ucl}(\sigma(\phi_i), \phi_i), \alpha \upharpoonright_i) \tag{4.9}$$

and  $S(\text{ucl}(\sigma(\phi_i \rightarrow \phi_j), \phi_i \rightarrow \phi_j), \alpha \upharpoonright_{\phi_i \rightarrow \phi_j})$ . Then the latter sentence is equivalent to

$$\forall \beta \sim_{\sigma(\phi_i \rightarrow \phi_j)} \alpha \upharpoonright_{\phi_i \rightarrow \phi_j} \ S(\phi_i, \beta \upharpoonright_i) \rightarrow S(\phi_j, \beta \upharpoonright_j)$$

which implies

$$\left( \forall \beta \sim_{\sigma(\phi_i \rightarrow \phi_j)} \alpha \upharpoonright_{\phi_i \rightarrow \phi_j} \ S(\phi_i, \beta \upharpoonright_i) \right) \rightarrow \left( \forall \beta \sim_{\sigma(\phi_i \rightarrow \phi_j)} \alpha \upharpoonright_{\phi_i \rightarrow \phi_j} \ S(\phi_j, \beta \upharpoonright_j) \right)$$

The above is equivalent to

$$\left( \forall \beta \sim_{\sigma(\phi_i)} \alpha \upharpoonright_{\phi_i} \ S(\phi_i, \beta \upharpoonright_i) \right) \rightarrow \left( \forall \beta \sim_{\sigma(\phi_j)} \alpha \upharpoonright_{\phi_j} \ S(\phi_j, \beta \upharpoonright_j) \right)$$

because by taking restriction in  $S(\phi_i, \beta \upharpoonright_i)$ , we ignore the irrelevant variables. The antecedent of the above is equivalent to (4.9) and this step finishes our whole proof.  $\square$

**Corollary 169** (Theorem 141).  $CT_0$  proves the Global Reflection Principle.

*Proof.* Working in  $CT_0$  let us put

$$S(\phi, \alpha) \equiv (\text{Form}_{\mathcal{L}_{\text{PA}}}(\phi) \wedge \alpha \in \text{Asn}(\phi) \wedge T(\phi[\alpha]))$$

Then, by Proposition 107,  $S(\phi, \alpha)$  is a formula of  $\mathcal{L}_T$  satisfying the axioms of  $CS_0$ . Hence by the above theorem for any sentence  $\phi$  we have

$$\text{Pr}_{\text{PA}}(\phi) \rightarrow \forall \alpha \in \text{Asn}(\phi) \ S(\phi, \alpha).$$

Since  $\phi$  is a sentence, then, in particular, we get

$$\text{Pr}_{\text{PA}}(\phi) \rightarrow S(\phi, \varepsilon),$$

and by the definition of  $S(\phi, \alpha)$

$$\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi)$$

which ends the proof.  $\square$

4.2 Many Faces of  $CT_0$ 

In this section we shall discuss various possible axiomatisations of  $CT_0$ . The result proved above will be used to complete the picture of interdependencies between various truth axioms studied earlier by (*inter alia*) Cieśliński and Enayat.

Now the promised "Many Faces" theorem:

**Theorem 170.** *The following theories have the same consequences:*

1.  $CT_0$
2.  $CT^- + \forall\phi (\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi))$
3.  $CT^- + \forall\phi (\text{Pr}_{\text{PA}}^T(\phi) \rightarrow T(\phi))$
4.  $CT^- + \forall\phi (\text{Pr}_{\emptyset}(\phi) \rightarrow T(\phi))$
5.  $CT^- + \forall\phi (\text{Pr}_{\text{CPC}}^T(\phi) \rightarrow T(\phi))$
6.  $CT^- + \text{DC} + \text{INT}$

*Proof.* (1)  $\Rightarrow$  (2) has been proved in the last section. The proofs of

$$(3) \Rightarrow (2)$$

$$(2) \Rightarrow (4)$$

$$(3) \Rightarrow (5)$$

follows directly from the definitions of respective theories. Proofs of the rest of implications require more subtle reasoning: (5)  $\Rightarrow$  (1) has been proved in [7], whereas (4)  $\Rightarrow$  (2) was shown in [7]. The implication (2)  $\Rightarrow$  (3) was established in [4]: we sketch the proof for Reader's convenience:

(2)  $\Rightarrow$  (3): We work in  $CT^- + \forall\phi(\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi))$ . Since all axioms of PA are true, what is left to show is that provability in First Order Logic preserves truth. Let  $\psi_0, \dots, \psi_x = \phi$  e a proof of  $\phi$  in First Order Logic such that for every  $i < x$  we have

$$T(\psi_i)$$

By Deduction Theorem ( $\bigwedge_{i < x} \psi_i \rightarrow \phi$ ) is provable in PA, hence (by the axiom for disjunction and negation and the Global Reflection Principle) we have

$$T\left(\bigwedge_{i < x} \psi_i\right) \rightarrow T(\phi)$$

In particular, it is enough to demonstrate  $T(\bigwedge_{i < x} \psi_i)$ . For  $j < x$  let us define

$$\begin{aligned} \theta'_j &= \bigwedge_{i \leq j} \psi_i \\ \theta_j(v) &= (v = \underline{j}) \rightarrow \theta'_j \\ \gamma(v) &= \bigwedge_{j < x} \theta_j(x) \end{aligned}$$

Let us make the following claim

**Claim 1.** For every  $j \leq x$  we have

$$T(\gamma(\underline{j})) \equiv T(\theta'_j)$$

*Proof of Claim 1.* Let us assume that  $j \leq x$ . For every  $j \neq i \leq x$  we have

$$\text{PA} \vdash \underline{i} \neq \underline{j}$$

Hence for any such  $i$

$$\text{PA} \vdash (\underline{i} = \underline{j}) \rightarrow \theta'_i$$

ad consequently

$$\text{PA} \vdash \bigwedge_{i \leq x, i \neq j} \theta_i \tag{4.10}$$

By the Global Reflection Principle we have  $T(\bigwedge_{i \leq x, i \neq j} \theta_i)$ . Moreover PA proves that whether conjunction holds does not depend on the way it is parenthesized, hence in particular, we have

$$\text{PA} \vdash \gamma(\underline{j}) \equiv \left( \left( \bigwedge_{i \leq x, i \neq j} \theta_i \right) \wedge \theta_j \right)$$

Hence by the above and 4.10 we get

$$\text{PA} \vdash \gamma(\underline{j}) \equiv \theta_j$$

Using the Global Reflection Principle and the compositional axioms of  $\text{CT}^-$  we get

$$T(\gamma(\underline{j})) \equiv T(\theta_j)$$

which ends the proof of the claim.  $\square$

Let us observe that  $\gamma(v)$  is a formula of  $\mathcal{L}_{\text{PA}}$ , hence we have

$$T(\text{ind}(v, \gamma))$$

and applying finitely many times compositional axioms we obtain

$$\left( T(\gamma(\underline{0})) \wedge \forall x (T(\gamma(\underline{x})) \rightarrow T(\gamma(\underline{x+1}))) \right) \rightarrow \forall x T(\gamma(\underline{x}))$$

By Claim 1 it is enough to demonstrate the antecedent of the above implication. Once again by the claim and our assumption on  $\psi_i$ ,  $T(\gamma(\underline{0}))$  holds. So let us fix arbitrary  $y$  and assume  $T(\gamma(\underline{y}))$  holds. If  $y+1 \geq x$ , then also  $T(\gamma(\underline{y+1}))$  holds for every  $i < x$

$$\text{PA} \vdash \underline{y+1} \neq \underline{i}$$

and, in particular,  $\text{PA} \vdash \theta_i(\underline{y+1})$  for every  $i < x$ . Hence  $\text{PA} \vdash \gamma(\underline{y+1})$ . So let assume that  $y+1 < x$ . By the claim it is enough to demonstrate  $T(\theta_{y+1})$ . By the Global Reflection Principle the  $T(\theta_{y+1})$  is equivalent to

$$T(\theta_y \wedge \psi_{y+1})$$

By our induction assumption and once again the claim we have  $T(\theta_y)$ . Since by our initial assumption  $T(\psi_{y+1})$  holds as well, application of compositional axioms in  $\text{CT}^-$  ends the proof.

Proof of (1)  $\Rightarrow$  (6) is a rather straightforward application of  $\Delta_0$  induction. Indeed, INT is provable in  $\text{CT}_0$ , since for every formula with at most one free variable  $\phi$

$$\theta(x) := T(\phi(\underline{x}))$$

is a  $\Delta_0$  formula of  $\mathcal{L}_T$  with parameter  $\phi$  (it involves some bounded quantifiers since we have to substitute  $\underline{x}$  to  $\phi$  for  $v$ ). Hence for every  $\phi$ ,

$$\left( T(\phi(\underline{0})) \wedge \forall x (T(\phi(\underline{x})) \rightarrow T(\phi(\underline{x+1}))) \right) \rightarrow \forall x T(\phi(\underline{x}))$$

is an axiom of  $\text{CT}_0$ . To demonstrate the Disjunctive Correctness axiom we use the relative truth definability of  $\text{CS}_0$  in  $\text{CT}_0$  (see Proposition 107) and the fact that  $\text{CS}_0 \vdash \text{DC}(S)$  (Proposition 166).

Proof of (6)  $\Rightarrow$  (1) is an unpublished result due to Ali Enayat: let us outline the proof. By Proposition 56 it is enough to demonstrate that in every model of  $\text{CT}^- + \text{DC} + \text{INT}$  (the extension of)  $T$  is a class. Let us fix arbitrary

$$\mathcal{M} \models \text{CT}^- + \text{DC} + \text{INT}$$

and  $c \in M$ . From now on we work in  $\mathcal{M}$ . Let  $\phi_0, \dots, \phi_a$  be any enumeration of sentences smaller than  $c$ . We define for  $i \leq a$

$$\begin{aligned} \psi_i(v) &:= (v = \underline{i}) \wedge \phi_i \\ \theta(v) &:= \bigvee_{i \leq a} \psi_i \end{aligned}$$

**Claim 2.** For every  $i \leq a$

$$T(\theta(\underline{i})) \equiv T(\phi_i)$$

*Proof of Claim 2.* Let us fix  $i \leq a$ . Let us observe that by the axioms of  $\text{CT}^-$ ,  $T(\phi_i)$  is equivalent to

$$T(\underline{i} = \underline{i} \wedge \phi_i)$$

Moreover if  $j \neq i$ , then we have  $\neg T(\underline{j} = \underline{i})$  and consequently

$$\neg T(\underline{j} = \underline{i} \wedge \phi_j)$$

Hence, by Disjunctive Correctness  $T(\theta(\underline{i}))$  is equivalent to  $T(\psi_i)$ , which in turn is equivalent to  $T(\phi_i)$ .  $\square$

Let us define:

$$\begin{aligned} \gamma(v, w_0) &:= (v \in w_0) \equiv \theta(v) \\ \zeta(w_1) &= \exists w_0 \forall v < w_1 (v \leq \underline{a} \rightarrow \gamma(v, w_0)) \end{aligned}$$

Intuitively,  $\zeta(w_1)$  says that there exists a set of whose elements are indices of true (in the sense of  $T$ ) sentences below  $c$ . Let us observe that, by compositional axioms of  $\text{CT}^-$ ,  $T(\zeta(\underline{b}))$  is equivalent to

$$\exists x \forall y < \min\{b, a + 1\} (y \in x \equiv T(\theta(\underline{y}))) \quad (4.11)$$

and by our claim:

$$\exists x \forall y < \min\{b, a + 1\} (y \in x \equiv T(\psi_y))$$

In particular,  $T(\zeta(\underline{a+1}))$  implies that there exists a code of all true sentences below  $c$ . By internal induction on  $z$  we shall show that  $\forall z T(\zeta(\underline{z}))$  holds. By 4.11,  $T(\zeta(\underline{0}))$  holds trivially. Let us fix arbitrary  $z$  and assume that  $T(\zeta(\underline{z}))$  holds. If  $z \geq a + 1$ , then, once again by 4.11  $T(\zeta(\underline{z+1}))$  holds as well. Assume that  $z \leq a$ . By our induction assumption there is a  $c$  such that

$$\forall y < z (y \in c \equiv T(\psi_y))$$

Let us fix  $c$ . Define

$$d = \begin{cases} c, & \text{if } \neg T(\psi_{z+1}) \\ c \cup \{z+1\}, & \text{if } T(\psi_{z+1}) \end{cases}$$

For such  $d$  we have

$$\forall y < z + 1 (y \in d \equiv T(\psi_y))$$

and the proof is finished. □

## 5. NON-CLASSICALLY COMPOSITIONAL TRUTH THEORIES

As we already saw, in the absence of induction, resignation from the axiom for the negation might result in weakening the theory:  $CT^-$  is model-theoretically non-conservative over PA whereas both  $PT^-$  and  $WPT^-$  are model-theoretically not stronger than PA itself. In the current chapter, we will investigate the strength are various reflection principles when added to non-classically compositional theories of truth. In particular, we shall revisit Theorem 170 and ask whether it is still true when considered with  $PT^-/WPT^-$  playing the role of  $CT^-$ . Moreover, we will be interested whether the distinction between *closure* reflection principles and *completeness* reflection principles becomes visible over non-classically compositional theories. We will start by considering the strength of  $PT^-$  and  $WPT^-$  extended with  $\Delta_0$  induction for  $\mathcal{L}_T$ .

### 5.1 Bounded induction

It is relatively easy to show that in the presence of  $\Sigma_1$  induction all the three theories, that we call  $CT_1$  ( $CS_1$ ),  $PT_1$  ( $PS_1$ ) and  $WPT_1$  ( $WPS_1$ ), are *the same* - by an obvious use of induction on the build-up for formulae (inside PA), one shows that non-classically compositional theories prove that each formula is total and consistent. We shall prove that it is *almost* true for variants of these theories equipped with  $\Delta_0$ - $\mathcal{L}_T$  induction only. We have stated "almost" since our result holds for  $CT_0$ ,  $PT_0$  and *an extension of*  $WPT_0$ , which we call  $WPT_0^{++}$ . This modification is, however, very natural and in our opinion is better crafted than one would normally call  $WPT_0$ . This change is motivated in the following way: in natural language, we usually express propositions of the form "All P's are also Q's", the famous example being

All men are mortal. (\*)

Then, when taking first lessons in First Order Logic, we learn that the above can be translated to the formal language as

$$\forall x(\text{Man}(x) \rightarrow \text{Mortal}(x))$$

or in the absence of the implication symbol

$$\forall x(\neg \text{Man}(x) \vee \text{Mortal}(x))$$

The last sentence literally says:

Every object in the world is either not a man or is mortal. (\*\*)

The translation of  $*$  into Classical First Order Logic, resulting in the sentence directly corresponding to  $**$ , rests upon creating a new complex property "being not a man or being mortal" and predicating it on every object in the world. There is, however, another way to render the meaning of  $*$ : we can see the quantifier "Every" as needing two properties (formulae) to form a sentence. The former property ("Being a man", in our example) restricts the range of quantification while the second one is predicated on every object from the new domain. Treating universal quantifiers as (possibly infinitary) conjunctions,  $*$  can be seen as

$$\bigwedge_{a: \text{Man}(a)} \text{Mortal}(a) \quad (\wedge *)$$

whereas  $**$  as

$$\bigwedge_a (\neg \text{Man}(a) \vee \text{Mortal}(a)) \quad (\wedge **)$$

Now, the difference between (possibly infinitary) sentences becomes more clearly visible on another example: consider the sentence "Every number less than 3 is prime"<sup>1</sup>. Analysed in the fashion of  $\wedge *$  it can be presented as

$$\text{Prime}(0) \wedge \text{Prime}(1) \wedge \text{Prime}(2)$$

which can be written using finitely many symbols. This is clearly not the case of the analogue of  $\wedge **$

$$(0 \geq 3 \vee \text{Prime}(0)) \wedge (1 \geq 3 \vee \text{Prime}(1)) \wedge (2 \geq 3 \vee \text{Prime}(2)) \wedge (3 \geq 3 \vee \text{Prime}(3)) \wedge \dots$$

Working in classical logic, we need not add those new binary quantifiers to the language, provided our aim is to describe the truth conditions for sentences such as  $*$ . However, if we would like our formalism to mimic the superficial form of natural language sentences more closely, there might still be good reasons to introduce these new means to the formal language. This line was pursued e.g. by Stephen Neale in [37] where he reconstructed Russell's theory of descriptions in this enriched logic.<sup>2</sup>

How do the above remarks transfer to the setting of axiomatic truth theories? We shall extend the arithmetised language with certain generalised quantifiers of the form  $\{\exists x : \phi\}\psi$ , where  $\phi$  is meant to restrict the range of quantification. The arithmetised language extended with those additional symbols will be denoted by  $\mathcal{L}_{\text{PA}}^{++}$ . Consequently,  $\text{Form}_{\mathcal{L}_{\text{PA}}^{++}}(x)$ ,  $\text{Sent}_{\mathcal{L}_{\text{PA}}^{++}}(x)$  etc. are arithmetical formulae strongly representing the set of (Gödel codes of) formulae, sentences etc. of  $\mathcal{L}_{\text{PA}}^{++}$ . We extend also our convention of using metavariables  $\phi, \psi, \dots, \phi(v), \psi(v), \dots$  to this new language; i.e. when dealing with  $\text{WPS}_0^{++}$  and  $\text{WPT}_0^{++}$ ,  $\forall \phi \dots$  is to be read as  $\forall x (\text{Form}_{\mathcal{L}_{\text{PA}}^{++}}(x) \rightarrow \dots)$ .

New quantifiers  $\{Qv : \phi\}\psi$  are to be treated as conjunctions (in the case of universal quantifier) or disjunctions (in the case of the existential one) of all sentences of the form

$$\psi(a)$$

<sup>1</sup> We did not promise mentioning only true sentences.

<sup>2</sup> One of the standard arguments against Russellian analysis is that what Russell called the logical form of sentences containing definite descriptions does not match their superficial subject-predicate form. Neale showed that if we allow for binary quantifiers (more than those discussed), then a more natural logical form of sentences with definite descriptions can be defined in a Russellian vein.

for any object  $a$  such that  $\phi(a)$ . In order for such a conjunction to be fully determined, we should know two things:

1. which sentences form the conjunction; i.e. what are the objects satisfying  $\phi$
2. whether  $\psi$  is well defined for those objects.

To formalise the second objective, let us define the generalised version of tot

$$\text{tot}_v(\phi, \psi) := \forall y \left( T(\phi(\underline{y}/v)) \rightarrow (T(\psi(\underline{y}/v)) \vee T(\neg\psi(\underline{y}/v))) \right)$$

Then the truth conditions for  $\{\exists x : \phi\}\psi$  (as usual,  $\{\forall x : \phi\}\psi$  will be defined as the dual) in Weak Kleene Logic should be:

$$\begin{aligned} \mathbf{G}\exists \forall v \forall \phi(v) \forall \psi(v) & \left( T(\{\exists v : \phi\}\psi) \equiv \text{tot}_v(\phi) \wedge \text{tot}_v(\phi, \psi) \wedge \exists x (T(\phi(\underline{x})) \wedge T(\psi(\underline{x}))) \right) \\ \mathbf{-G}\exists \forall v \forall \phi(v) \forall \psi(v) & \left( T(\neg\{\exists v : \phi\}\psi) \equiv \text{tot}_v(\phi) \wedge \forall x (T(\phi(\underline{x})) \rightarrow T(\neg\psi(\underline{x}/v))) \right) \end{aligned}$$

Let  $\text{WPT}_0^{++}$  denote the theory containing  $\mathcal{L}_{\text{PA}}^{++}$ -variants of axioms of  $\text{WPT}^-$  (Definition 96). In the context of  $\text{WPT}_0^{++}$  we always unravel  $\forall \phi(\bar{x})\Phi(\phi)$  as  $\forall x(\text{Form}_{\mathcal{L}_{\text{PA}}^{++}}(x) \rightarrow \Phi(x))$ . We shall treat  $\{\forall v : \phi\}\psi$  as the abbreviation of

$$\neg\{\exists v : \phi\}\neg\psi$$

for arbitrary formulae  $\phi, \psi$  (this will be needed later on when introducing  $\text{WPS}_0^{++}$ ). With such a definition, we have:

**Proposition 171.** *The following conditions are provable in  $\text{WPT}_0^{++}$ :*

1.  $\forall v \forall \phi(v) \forall \psi(v) \left( T(\{\forall v : \phi\}\psi) \equiv \text{tot}_v(\phi) \wedge \forall x (T(\phi(\underline{x})) \rightarrow T(\psi(\underline{x}/v))) \right)$
2.  $\forall v \forall \phi(v) \forall \psi(v) \left( T(\neg\{\forall v : \phi\}\psi(v)) \equiv \text{tot}_v(\phi) \wedge \text{tot}_v(\phi, \psi) \wedge \exists x (T(\phi(\underline{x})) \wedge T(\neg\psi(\underline{x}))) \right)$

Then we can prove the following theorems

**Theorem 172.**  $\text{PT}_0 \vdash \forall \phi \left( T(\ulcorner \neg\phi \urcorner) \equiv \neg T(\phi) \right)$

**Theorem 173.**  $\text{WPT}_0^{++} \vdash \forall \phi \left( T(\ulcorner \neg\phi \urcorner) \equiv \neg T(\phi) \right)$

**Remark 174.** To compare theories formulated over various signatures, it will be useful to introduce a notion that would allow us to say when two such theories are *the same, up to the translation of their logical symbols*. For example, given Theorem 173, we would like to say that  $\text{WPT}_0^{++}$  and  $\text{CT}_0$  are the same theories. This is not literally true, since  $\text{CT}_0$  does not contain axioms governing the use of new symbols dealt with in  $\text{WPT}_0^{++}$ . However,  $\text{CT}_0$  can define the appropriate semantics for those new symbols; hence the difference between the two theories is purely *notational*. In the definition, below, we shall use the notion of a *translation* introduced in Definition 52.

**Definition 175.** Let  $\text{Th}_1$  and  $\text{Th}_2$  be two truth theories with truth predicates  $T_1$  and  $T_2$  and let  $\mathcal{L}_{\text{PA}}^1$  and  $\mathcal{L}_{\text{PA}}^2$  be two languages over the same set of variables  $\text{Var}$  and non-logical constants  $\lambda = \{0, 1, +, \cdot\}$  (see Definition 50). Hence,  $\mathcal{L}_{\text{PA}}^1$  and  $\mathcal{L}_{\text{PA}}^2$  differ at most on propositional connectives and quantifiers present in both languages. Assume that  $\text{Th}_1$  and  $\text{Th}_2$  define the truth conditions for  $\mathcal{L}_{\text{PA}}^1$  and  $\mathcal{L}_{\text{PA}}^2$  respectively. Moreover let  $\mathcal{L}$  be a language. We shall say that  $\text{Th}_2$  is a *mod*  $\mathcal{L}$  translational subtheory of  $\text{Th}_1$  if there exists a PA provable  $\mathcal{L}$ -conservative translation  $*$  such that for every axiom  $\Theta$  of  $\text{Th}_2$

$$\text{Th}_1 \vdash \Theta[T_1(t^*)/T_2(t)]$$

where (using Convention 2)  $T_1(t^*)$  abbreviates  $\exists x(x = t^* \wedge T_1(x))$  (where  $x$  is some variable not appearing in  $t$ ) and  $z = t^*$  is an arithmetical formula representing in PA the translation function. In other words we might say that the formula  $T_1(t^*)$  relatively truth defines  $\text{Th}_2$  in  $\text{Th}_1$ . We shall say that  $\text{Th}_2$  and  $\text{Th}_1$  are *mod*  $\mathcal{L}$  mutual notational variants if  $\text{Th}_1$  is a *mod*  $\mathcal{L}$  translational subtheory of  $\text{Th}_2$  and vice versa.

The above definition is analogous for theories of satisfaction.

Let us observe that the above definition would be superfluous if we assumed that all the truth (satisfaction) theories we consider are formulated over the same arithmetical language and treat the additional symbols as simply internal abbreviations of certain constructions. For example we could have stipulated that  $\mathcal{L}_{\text{PA}}$  contains the newly introduced quantifiers and in  $\text{CT}^-$  we have the following axiom at our disposal:

$$\forall v \forall \phi(v) \forall \psi(v) \quad (T(\{\forall v : \phi\}\psi) \equiv T(\forall v(\phi \rightarrow \psi)))$$

Under such an assumption, we could have obtained the conclusion that all theories considered by us are *equal*. This solution is, however, a little bit artificial and we find it less elegant. The current approach makes the relation between theories considered more explicit and reduces the number of *ad hoc* assumptions.

**Remark 176.**  $\text{WPT}_0^{++}$  is a *mod*  $\mathcal{L}_{\text{PA}}$  translational subtheory of  $\text{CT}_0$ . Indeed, we can define the translation  $*$  by putting

$$\begin{aligned} \forall \phi \forall \psi \forall v \quad ((\{\forall v : \phi\}\psi)^* &= \forall v(\phi^* \rightarrow \psi^*)) \\ \forall \phi \forall \psi \forall v \quad ((\{\exists v : \phi\}\psi)^* &= \exists v(\phi^* \wedge \psi^*)) \end{aligned}$$

Then for example  $\text{CT}_0$  proves

$$\forall v \forall \phi(v) \forall \psi(v) \quad T((\{\forall v : \phi\}\psi)^*) \equiv \left( \text{tot}_v(\phi^*) \wedge \forall x(T(\phi(x)^*) \rightarrow T(\psi(x)^*)) \right)$$

As we already announced, we will prove both Theorem 172 and Theorem 173 for versions of two theories with the satisfaction predicate. Let us first formulate the appropriate extension of  $\text{WPS}_0$ . As in the case of  $\text{WPT}_0^{++}$  it contains  $\mathcal{L}_{\text{PA}}^{++}$  – variants of  $\text{WPS}^-$  axioms. To introduce two specific axioms of  $\text{WPS}_0^{++}$ , let us generalise restricted totality to the language with satisfaction predicate:

**Definition 177.**

$$\text{tot}_v(\phi, \psi, \alpha) := \forall \beta \sim_v \alpha \left( S(\phi, \beta \upharpoonright_\phi) \rightarrow (S(\psi, \beta \upharpoonright_\psi) \vee S(\neg\psi, \beta \upharpoonright_\psi)) \right)$$

And now two additional axioms of  $\text{WPS}_0^{++}$ :

$$\mathbf{G}\exists_{\mathbf{S}} \forall \phi(\bar{w}) \forall \psi(\bar{w}) \forall \alpha \in \text{Asn}(\phi, \psi) \forall v$$

$$\left( S(\{\exists v : \phi\}\psi, \alpha) \equiv \text{tot}_v(\phi, \alpha \upharpoonright_\phi) \wedge \text{tot}_v(\phi, \psi, \alpha) \wedge \exists \beta \sim_v \alpha (S(\phi, \beta \upharpoonright_\phi) \wedge S(\psi, \beta \upharpoonright_\psi)) \right)$$

$$\neg\mathbf{G}\exists_{\mathbf{S}} \forall \phi(\bar{w}) \forall \psi(\bar{w}) \forall \alpha \in \text{Asn}(\phi(\bar{w}), \psi(\bar{w})) \forall v$$

$$\left( S(\{\neg\exists v : \phi\}\psi, \alpha) \equiv \text{tot}_v(\phi, \alpha \upharpoonright_\phi) \wedge \forall \beta \sim_v \alpha (S(\phi, \beta \upharpoonright_\phi) \rightarrow S(\neg\psi, \beta \upharpoonright_\psi)) \right)$$

Similarly to  $\text{WPT}_0^{++}$  case, we have the following proposition (we remind that quantifiers  $\forall \phi(\bar{x})$  range over formulae of  $\mathcal{L}_{\text{PA}}^{++}$ ):

**Proposition 178.** *The following sentences are provable in  $\text{WPS}_0^{++}$*

$$1. \forall \phi(\bar{w}) \forall \psi(\bar{w}) \forall \alpha \in \text{Asn}(\phi, \psi) \forall v$$

$$\left( S(\{\forall v : \phi\}\psi, \alpha) \equiv \text{tot}_v(\phi, \alpha \upharpoonright_\phi) \wedge \forall \beta \sim_v \alpha (S(\phi, \beta \upharpoonright_\phi) \rightarrow S(\psi, \beta \upharpoonright_\psi)) \right)$$

$$2. \forall \phi(\bar{w}) \forall \psi(\bar{w}) \forall \alpha \in \text{Asn}(\phi, \psi) \forall v$$

$$\left( S(\{\neg\forall v : \phi\}\psi, \alpha) \equiv \text{tot}_v(\phi, \alpha \upharpoonright_\phi) \wedge \text{tot}_v(\psi, \phi, \alpha) \wedge \exists \beta \sim_v \alpha (S(\phi, \beta \upharpoonright_\phi) \wedge S(\neg\psi, \beta \upharpoonright_\psi)) \right)$$

Now we are about to prove (recall that, by our convention,  $\forall \phi(\bar{w})$  in the first of the theorems below has different meaning than  $\forall \phi(\bar{w})$  in the second one).

**Theorem 179.**  $\text{PS}_0 \vdash \forall \phi(\bar{w}) \forall \alpha \in \text{Asn}(\phi) (S(\neg\phi, \alpha) \equiv \neg S(\phi, \alpha))$

**Theorem 180.**  $\text{WPS}_0^{++} \vdash \forall \phi(\bar{w}) \forall \alpha \in \text{Asn}(\phi) (S(\neg\phi, \alpha) \equiv \neg S(\phi, \alpha))$

**Remark 181.** We do not know whether extending  $\text{WPS}_0$  is necessary for proving that all formulae are total and consistent. However, as we shall demonstrate soon, This aim is accomplished by a much weaker than  $\text{WPS}_0^{++}$  extension of  $\text{WPS}_0$ .

Let us show how theorems 172 and 173 follow from the two above:

*Proof of Theorem 172 assuming Theorem 179.* Working in  $\text{PT}_0$  let us define

$$S(\phi, \alpha) := \text{Form}_{\mathcal{L}_{\text{PA}}}(\phi) \wedge \alpha \in \text{Asn}(\phi) \wedge T(\phi[\alpha]) \quad (5.1)$$

Then, by Proposition 107, the axioms of  $\text{PS}_0$  are provable in  $\text{PT}_0$  with so defined satisfaction predicate. By Theorem 179 this satisfaction predicate satisfies the axioms of  $\text{CS}_0$ . If  $\phi$  is an arbitrary sentence, then the following are equivalent ( $\varepsilon$  denotes the empty function):

1.  $T(\neg\phi)$
2.  $T(\neg\phi[\varepsilon])$
3.  $S(\neg\phi, \varepsilon)$
4.  $\neg S(\phi, \varepsilon)$
5.  $\neg T(\phi[\varepsilon])$
6.  $\neg T(\phi)$

□

Proof of Theorem 173, assuming Theorem 180, is identical. Before proving Theorem 179, let us outline what is problematic in working with  $\Delta_0$  induction only (we shall make precise the remarks stated in Section 3.3.3): the routine proof that every formula is total and consistent would use induction on the build-up of formulae. But to secure the induction step for the existential quantifier, i.e. to show that

$$S(\neg\exists v\phi, \alpha) \equiv \neg S(\exists v\phi, \alpha)$$

we should know that  $S(\neg\phi, \beta) \equiv \neg S(\phi, \beta)$  holds for arbitrary  $\beta$  differing from  $\alpha$  at most on  $v$ . In this assumption however the size of  $\beta$  cannot be bounded; hence this formula is at least  $\Pi_1(\text{PAT})$ . We encountered the same problem when proving the Global Reflection Principle in  $\text{CT}_0$  and we will show that the same technique of reducing the complexity of the induction assumption can be used also in the present context.

*Proof of Theorem 179.* As observed in Corollary 149  $\text{PS}_0$  proves the commutativity with blocks of universal quantifiers. Working in  $\text{PS}_0$  let us fix an arithmetical formula  $\phi$  and an assignment  $\alpha \in \text{Asn}(\phi)$ . For  $\pi \in \text{Skel}(\phi)$  let  $\sigma(\pi)$  be the sequence define in Definition 159. Let  $k$  be the complexity of  $\phi$ . Let  $l(\pi)$  abbreviate  $l^\phi(\pi)$ . We first show that  $\phi$  is total, and then, via a dual argument, that it is consistent. For the former part: by induction on  $n$  we show

$$\underbrace{\forall n \forall \pi \in \text{Skel}(\phi) \left( \text{len}(\pi) \geq k - n \longrightarrow S(\text{ucl}(\sigma(\pi), l(\pi) \vee \neg l(\pi)), \alpha \upharpoonright \cdot) \right)}_{:= \Psi(y)} \quad (*)$$

The above, for  $n > k$  will imply that we have  $S(\phi \vee \neg\phi, \alpha)$  and hence, since  $\phi$  and  $\alpha$  were arbitrary,  $\text{Tot}(S)$ . The use of induction on  $n$  in  $\Psi(n)$  is legitimate in  $\text{PS}_0$  since by collection axioms in PA there exists a  $u$  such that

$$\forall \pi \in \text{Skel}(\phi) \text{ ucl}(\sigma(\pi), l(\pi) \vee \neg l(\pi)) < u$$

hence we can apply Remark 57 to justify the use of induction. The base step for  $n = 0$  is obvious, since atomic formulae are total provably in  $\text{PS}^-$  and, by Corollary 149 this suffices to complete the first step of induction. So let us fix  $n$  and assume that  $\Psi(n)$  holds. Let us fix arbitrary  $\pi$  of length  $k - (n + 1)$  and let  $\theta = l(\pi)$ . By the axiom for double negation in  $\text{PT}^-$  the case when  $\theta = \neg\theta'$  for some  $\theta'$  can be reduced to the step for  $\theta'$ . We shall show the steps for  $\vee$  and  $\exists$ .

Suppose first that  $\theta = \theta_1 \vee \theta_2$ . Then by the commutativity with blocks of universal quantifiers we have:

$$S(\text{ucl}(\sigma(\pi), \theta \vee \neg\theta), \alpha \uparrow) \equiv \forall \beta \sim_{\sigma(\pi)} \alpha \uparrow_{\theta \vee \neg\theta} S(\theta \vee \neg\theta, \beta \uparrow)$$

Let us fix  $\beta \sim_{\sigma(\theta)} \alpha \uparrow_{\theta \vee \neg\theta}$  and focus on  $S(\theta \vee \neg\theta, \beta \uparrow)$ . We have:

$$\begin{aligned} S(\theta \vee \neg\theta, \beta \uparrow) &\equiv S(\theta, \beta \uparrow) \vee S(\neg\theta, \beta \uparrow) \\ &\equiv S(\theta_1, \beta \uparrow) \vee S(\theta_2, \beta \uparrow) \vee (S(\neg\theta_1, \beta \uparrow) \wedge S(\neg\theta_2, \beta \uparrow)) \end{aligned}$$

The last sentence is implied by

$$S(\theta_1 \vee \neg\theta_1, \beta \uparrow) \wedge S(\theta_2 \vee \neg\theta_2, \beta \uparrow)$$

which is true by our inductive assumption and the fact that  $\text{im}(\sigma(\pi)) \cap \text{FV}(\theta_1) = \text{im}(\sigma(\pi \frown 0))$  and  $\text{im}(\sigma(\pi)) \cap \text{FV}(\theta_2) = \text{im}(\sigma(\pi \frown 1))$ .

Let us now suppose that  $\theta = \exists v \theta_1$ . Then  $\text{im}(\sigma(\pi \frown 0)) = \text{im}(\sigma(\pi)) \cup \{v\}$ . Let us fix arbitrary  $\beta \sim_{\sigma(\theta)} \alpha \uparrow_{\theta \vee \neg\theta}$ . Then

$$S(\theta \vee \neg\theta, \beta \uparrow) \equiv S(\theta, \beta \uparrow) \vee S(\neg\theta, \beta \uparrow)$$

The right-hand side is equivalent to

$$\exists \gamma \sim_v \beta \uparrow_{\theta} S(\theta_1, \gamma \uparrow) \vee \forall \gamma \sim_v \beta \uparrow_{\neg\theta} S(\neg\theta_1, \gamma \uparrow) \quad (*)$$

Let us observe that  $\gamma \sim_v \beta$  and  $\beta \sim_{\sigma(\theta)} \alpha$  hold jointly if and only if

$$\gamma \sim_{\sigma(\theta')} \alpha$$

holds. In particular, our inductive assumption implies that

$$\forall \beta \sim_{\sigma(\theta)} \alpha \forall \gamma \sim_v \beta (S(\theta_1, \gamma \uparrow) \vee S(\neg\theta_1, \gamma \uparrow)),$$

which, in turn, clearly implies \*.

Let us now turn to the other part of our proof. We use the same technique and by induction on  $n$  show that

$$\forall n \forall \pi \in \text{Skel} \left( \text{len}(\pi) \geq k - n \longrightarrow \neg S(\text{ecl}(\sigma(\pi), l(\pi) \wedge \neg l(\pi)), \alpha \uparrow) \right)$$

That we are allowed to use induction can be justified in the same way as above. Let us show the induction step for  $\vee$  and  $\exists$ . Fix arbitrary  $\pi \in \text{Skel}$  of length  $n - (k + 1)$  and let  $\theta = l(\pi)$ . By Corollary 150 we have

$$\neg S(\text{ecl}(\sigma(\pi), \theta \wedge \neg\theta), \alpha \uparrow) \equiv \neg \exists \beta \sim_{\sigma(\pi)} \alpha \uparrow_{\theta \wedge \neg\theta} S(\theta \wedge \neg\theta, \beta \uparrow)$$

Assume that  $\theta = \theta_0 \vee \theta_1$ . Let us observe that for every  $\beta \sim_{\sigma(\pi)} \alpha \uparrow_{\theta \wedge \neg\theta}$  we have

$$S(\theta \wedge \neg\theta, \beta \uparrow) \equiv (S(\theta_0, \beta \uparrow) \vee S(\theta_1, \beta \uparrow)) \wedge (S(\neg\theta_0, \beta \uparrow) \wedge S(\neg\theta_1, \beta \uparrow))$$

The sentence on the right implies

$$S(\theta_0 \wedge \neg\theta_0, \beta \uparrow) \vee S(\theta_1 \wedge \neg\theta_1, \beta \uparrow)$$

Hence, by induction assumption for  $\theta_0$  and  $\theta_1$  there can be no  $\beta \sim_{\sigma(\pi)} \alpha \upharpoonright_{\theta}$  such that  $S(\theta \wedge \neg\theta, \beta)$ . Let us now turn to  $\exists$  case: assume  $\theta = \exists v\theta_0$ . For every  $\beta \sim_{\sigma(\pi)} \alpha \upharpoonright_{\theta \wedge \neg\theta}$  we have

$$S(\theta \wedge \neg\theta, \beta \upharpoonright_{\cdot}) \equiv \exists \gamma \sim_v \beta S(\theta_0, \gamma \upharpoonright_{\cdot}) \wedge \forall \gamma \sim_v \beta S(\neg\theta_0, \gamma \upharpoonright_{\cdot})$$

The right-hand side of the above implies

$$\exists \gamma \sim_v \beta (S(\theta_0, \gamma \upharpoonright_{\cdot}) \wedge S(\neg\theta_0, \gamma \upharpoonright_{\cdot}))$$

which is impossible by the induction assumption on  $\theta_0$  ( $\pi \frown 0$  to be precise). This ends the whole proof.  $\square$

Let us now regard the  $\text{WPS}_0^{++}$  case. The proof strategy is essentially the same as in the case of  $\text{PS}_0$ , although we must indicate how to reprove Theorem 144 in the new setting. As stated in Remark 148, to reconstruct our proof of Theorem 144 in an axiomatic theory of satisfaction  $\text{Th}$ , it is sufficient that  $\text{Th}$  proves:

1.  $\forall \phi(\bar{w}) \forall v \forall \alpha \in \text{Asn}(\forall v \phi) \left( S(\forall v \phi, \alpha) \equiv \forall \beta \sim_v \alpha S(\phi, \beta \upharpoonright_{\phi}) \right)$
2.  $\forall \phi(\bar{w}) \forall v \forall \alpha \in \text{Asn}(\forall v \phi) \forall y \left( S(\forall v < \underline{y} \phi, \alpha) \equiv \forall \beta \sim_v \alpha (\beta(v) < y \rightarrow S(\phi, \beta \upharpoonright_{\phi})) \right)$
3.  $\Delta_0$  induction for  $\mathcal{L}_T$ .

The first one is an axiom of  $\text{WPS}^-$  and the third is, by definition, admissible in  $\text{WPS}_0$ . However, the second one fails in  $\text{WPS}^-$  since in the arithmetised language  $\forall v < \underline{y} \phi$  is a short for

$$\forall v (\neg(v < \underline{y}) \vee \phi)$$

And hence in  $\text{WPS}^-$   $S(\forall v < \underline{y} \phi, \alpha)$  implies

$$\text{tot}_v(\phi, \alpha).$$

The above sentence obviously is not, in general, a consequence of

$$\forall \beta \sim_v \alpha (\beta(v) < y \rightarrow S(\phi, \beta))$$

which can be easily seen by taking  $y = 0$  and  $\phi$  to be arbitrary formula which is not total with respect to  $v$ . In  $\text{WPS}_0^{++}$  we can remedy this with our notion of generalised quantifiers. For technical purposes it will be convenient to formulate next proposition and lemmata in greater generality. Let  $\mathcal{L}_{\text{PA}}^+$  denote the extension of arithmetised language with *bounded existential quantifier*; i.e. in this language we are allowed to build expressions of the form

$$\{\exists v : v < t\} \phi$$

where  $t$  is a term. Variant of  $\text{WPS}^-$  defined for this language will be denoted with  $\text{WPS}_+^-$

**Remark 182.** Warning: extension of this theory with  $\Delta_0$  induction will be denoted by  $\text{WPS}_0^+$ : unfortunately defining non-inductive theories with superscript "-" is almost as well-established (e.g.  $\text{PA}^-$ ) as writing subscript "0" to denote their extensions with  $\Delta_0$  induction.

In context of this theory, we interpret  $\forall\phi(\bar{x})$  as ranging over formulae of  $\mathcal{L}_{\text{PA}}^+$ ; i.e.  $\forall\phi(\bar{x})\Phi(\phi)$  is to be unravelled as

$$\forall x(\text{Form}_{\mathcal{L}_{\text{PA}}^+}(x) \rightarrow \Phi(x)).$$

$\text{WPS}_+^-$  contains  $\mathcal{L}_{\text{PA}}^+$ -variants of axioms of  $\text{WPS}^-$  and additionally the  $\mathcal{L}_{\text{PA}}^+$ -variants of the two following axioms:

$$\mathbf{G_b}\exists_{\mathbf{S}} \forall\phi(\bar{w})\forall v\forall t(\bar{w})\forall\alpha \in \text{Asn}(\phi, t)$$

$$(S(\{\exists v : v < t\}\phi, \alpha) \equiv \text{tot}_v(v < t, \phi, \alpha) \wedge \exists\beta \sim_v \alpha(\beta(v) < (t)_{\beta}^{\circ} \wedge S(\phi, \beta \uparrow.))$$

$$\neg\mathbf{G_b}\exists_{\mathbf{S}} \forall\phi(\bar{w})\forall v\forall t(\bar{w})\forall\alpha \in \text{Asn}(\phi, t)$$

$$(S(\neg\{\exists v : v < t\}\phi, \alpha) \equiv \forall\beta \sim_v \alpha(\beta(v) < (t)_{\beta}^{\circ} \rightarrow S(\neg\phi, \beta \uparrow.))$$

As usual we take

$$\{\forall v : v < t\}\phi$$

to be an abbreviation of

$$\neg\{\exists v : v < t\}\neg\phi$$

We can derive in  $\text{WPS}_+^-$  the natural truth conditions for  $\{\forall v : v < t\}\phi$  and  $\neg\{\forall v : v < t\}\phi$ , i.e. we have

**Proposition 183.**  $\text{WPS}_+^-$  proves

1.  $\forall\phi(\bar{w})\forall v\forall t(\bar{w})\forall\alpha \in \text{Asn}(\phi, t)(S(\{\forall v : v < t\}\phi, \alpha) \equiv \forall\beta \sim_v \alpha(\beta(v) < (t)_{\beta}^{\circ} \rightarrow S(\phi, \beta \uparrow.))$
2.  $\forall\phi(\bar{w})\forall v\forall t(\bar{w})\forall\alpha \in \text{Asn}(\phi, t)$

$$(S(\neg\{\forall v : v < t\}\phi, \alpha) \equiv \text{tot}_v(v < t, \phi, \alpha) \wedge \exists\beta \sim_v \alpha(\beta(v) < (t)_{\beta}^{\circ} \wedge S(\neg\phi, \beta \uparrow.))$$

Provability of the above is really what we needed. Let us show that it is a consequence of  $\text{WPS}_0^{++}$ :

**Proposition 184.**  $\text{WPS}_0^{++} \vdash \text{WPS}_+^-$

*Proof.* The statement follows by the  $\mathbf{G}\exists_{\mathbf{S}}$  and  $\neg\mathbf{G}\exists_{\mathbf{S}}$  axioms and the fact that for every variable  $v$  and term  $t$

$$v < t$$

is a formula of standard complexity, so

$$\forall\beta \in \text{Asn}(t) (\text{tot}_v(v < t, \beta))$$

is provable in  $\text{WPS}^-$ . □

With the above proposition in hand, we can prove an analogue of Lemma 145. To do this we have to adapt  $\text{bucl}(\cdot, \cdot, \cdot)$  function to the current arithmetised language. For a sequence of variables  $\sigma$ , formula  $\phi$  and number  $y$  let us put

$$\text{bucl}^+(\sigma, \phi, y) := \ulcorner \{\forall \sigma(0) : \sigma(0) < \underline{y}\} \{\forall \sigma(1) : \sigma(1) < \underline{y}\} \dots \{\forall \sigma(a) : \sigma(a) < \underline{y}\} \phi \urcorner$$

where  $y = \max(\text{dom}(\sigma))$ . Now, the following lemma follows by exactly the same proof as Lemma 145:

**Lemma 185.** *The following sentence is provable in  $\text{WPS}_0^+$ :*

$$\forall \phi(\bar{w}) \forall \sigma \forall v \in \text{Var} \setminus \text{Var}(\text{ucl}(\sigma, \phi)) \forall \alpha \in \text{Asn}(\text{bucl}^+(\sigma, \phi, v)) \\ \left( S(\text{bucl}^+(\sigma, \phi, v), \alpha) \equiv \forall \beta \left[ (\beta \sim_\sigma \alpha \wedge \beta \preceq \alpha[\sigma \mapsto (\alpha(v) - 1)]) \longrightarrow S(\phi, \beta \upharpoonright_\phi) \right] \right)$$

Similarly we can establish the analogues of Lemmata 146 and 147:

**Lemma 186.** *The following sentence is provable in  $\text{WPS}_0^+$ :*

$$\forall \phi(\bar{w}) \forall \sigma \forall v \in \text{Var} \setminus \text{Var}(\text{ucl}(\sigma, \phi)) \forall \alpha \in \text{Asn}(\text{ucl}(\sigma, \phi)) \\ \left( S(\forall v \text{bucl}^+(\sigma, \phi, v), \alpha) \equiv \forall \beta \sim_\sigma \alpha \ S(\phi, \beta \upharpoonright_\phi) \right)$$

**Lemma 187.** *The following sentence is provable in  $\text{WPS}_0^+$ :*

$$\forall \phi(\bar{w}) \forall \sigma \forall v \in \text{Var} \setminus \text{Var}(\text{ucl}(\sigma, \phi)) \forall \alpha \in \text{Asn}(\text{ucl}(\sigma, \phi)) \\ \left( S(\text{ucl}(\sigma, \phi), \alpha) \equiv S(\forall v \text{bucl}^+(\sigma, \phi, v), \alpha) \right)$$

By Proposition 105, point 2. and Theorem 179 the proof of Theorem 180 will be finished as soon as we demonstrate the following

**Lemma 188.**  $\text{WPS}_0^+ \vdash \text{Tot}(S)$ .

*Proof.* We adapt the same strategy as in the proof of Theorem 179. We work in  $\text{WPS}_0^+$  and fix a formula  $\phi$  and  $\alpha \in \text{Asn}(\phi)$ . Let  $k$  be the complexity of  $\phi$ . We use  $\Delta_0$  induction on  $n$  to demonstrate that

$$\underbrace{\forall n \forall \pi \in \text{Skel}(\phi) \left( \text{len}(\pi) \geq k - n \longrightarrow S(\text{ucl}(\sigma(\pi), l(\pi) \vee \neg l(\pi)), \alpha \upharpoonright_\cdot) \right)}_{:= \Psi(n)}$$

where the meanings of  $\sigma(\pi)$  and  $l(\pi)$  are the same as in the proof of Theorem 179. By the three above lemmata, the base step for  $n = 0$  is trivial since each atomic formula is total provably in  $\text{WPS}_+^-$ . To verify the induction step let us fix  $\pi$  of length  $n - (k + 1)$  and assume that  $\Psi(k)$  holds. Let  $\theta = l(\pi)$ . If  $\theta$  starts with the negation sign, then the proof easily reduces to induction assumption (by the double negation axiom in  $\text{WPS}^-$ ; as previously we have to use the three

preceding lemmata to "pull" the quantifier prefix outside  $S$ ). Let us show the steps for  $\forall, \exists$  and  $\{\exists v : v < w\}$ . By generalised commutativity with universal quantifier we have

$$S(\text{ucl}(\sigma(\pi), \theta \vee \neg\theta), \alpha \uparrow) \equiv \forall \beta \sim_{\sigma(\pi)} \alpha \uparrow_{\theta \vee \neg\theta} S(\theta \vee \neg\theta, \beta)$$

Let us fix arbitrary  $\beta \sim_{\sigma(\pi)} \alpha \uparrow_{\theta \vee \neg\theta}$ . Let us abbreviate  $S(\phi, \beta \uparrow)$  with  $S_\beta(\phi)$ . Then, using the axioms for disjunction and double negation in  $\text{WPS}^-$ , we have:

$$\begin{aligned} S(\theta \vee \neg\theta, \beta \uparrow) &\equiv (S_\beta(\theta) \wedge S_\beta(\neg\theta)) \vee S_\beta(\neg\theta) \vee S_\beta(\theta) \\ &\equiv S_\beta(\theta) \vee S_\beta(\neg\theta) \end{aligned}$$

Now we distinguish cases: assume that  $\theta = \theta_0 \vee \theta_1$ . Then, using the axiom for disjunction and for the negation of disjunction we see that the last sentence is equivalent to

$$\text{tot}(\theta_0, \beta) \wedge \text{tot}(\theta_1, \beta)$$

which clearly follows from our assumption on  $\theta_0$  and  $\theta_1$  ( $\pi \frown 0$  and  $\pi \frown 1$  being precise; once again we use the fact that  $\text{im}(\sigma(\pi)) \cap \text{FV}(\theta_0) = \text{im}(\sigma(\pi \frown 0))$  and  $\text{im}(\sigma(\pi)) \cap \text{FV}(\theta_1) = \text{im}(\sigma(\pi \frown 1))$ ). Let us now turn to the step for  $\exists$ . Suppose  $\theta = \exists v \theta_0$ . By the axiom for the existential quantifier and the negation of existential quantifier we get

$$\text{tot}_v(\theta_0, \beta) \wedge \left( (\exists \gamma \sim_v \beta S_\gamma(\theta_0)) \vee (\forall \gamma \sim_v \beta S_\gamma(\neg\theta_0)) \right)$$

The above is clearly equivalent to simply  $\text{tot}_v(\theta_0, \beta)$ , which in turn is equivalent to our induction assumption on  $\theta_0$ . Indeed, either  $v$  has a free occurrence in  $\theta_0$ , and in such situation  $v$  is listed in  $\sigma(\theta_0)$ , or not, and in such situation

$$\text{tot}_v(\theta_0, \beta) \equiv (S(\theta_0, \beta \uparrow) \vee S(\neg\theta_0, \beta \uparrow))$$

which follows from our induction assumption, since  $\sigma(\pi) = \sigma(\pi \frown 0)$ . Let us now assume that  $\theta = \{\exists v : v < t\} \theta_0$ . Then by the specific axioms of  $\text{WPS}_+^-$  we have

$$\forall \gamma \sim_v \beta (\gamma(v) < (t)_\beta^\circ \rightarrow S_\gamma(\neg\theta_0)) \vee \left( \text{tot}_v(v < t, \theta_0, \beta) \wedge \exists \gamma \sim_v \beta (\gamma(v) < (t)_\beta^\circ \wedge S_\gamma(\theta_0)) \right)$$

The above is implied by  $\text{tot}_v(v < t, \theta_0, \beta)$  i.e.

$$\forall \gamma \sim_v \beta (\gamma(v) < (t)_\beta^\circ \rightarrow (S_\gamma(\theta_0) \vee S_\gamma(\neg\theta_0)))$$

As we already justified while dealing with  $\exists$  case, our induction assumption on  $\theta_0$  ( $\pi \frown 0$ ) implies  $\text{tot}_v(\theta_0, \beta)$  and the latter in turn clearly implies the above restricted version of totality of  $\theta_0$ .  $\square$

The above proposition completes the proof of Theorem 180. Let us summarize our findings:

**Theorem 189.**

1. Theories  $\text{CT}_0$  and  $\text{PT}_0$  have the same consequences. The same holds for  $\text{CS}_0$  and  $\text{PS}_0$ .
2. Theories  $\text{WPT}_0^{++}$ ,  $\text{PT}_0$ ,  $\text{CT}_0$  are mutual notational variants of each other. The same holds for theories  $\text{WPS}_0^{++}$ ,  $\text{PS}_0$  and  $\text{CS}_0$ .

Let us now justify the introduction of  $\text{WPS}_0^+$ : it is a translational subtheory of another extension of  $\text{WPS}^-$  modelled after *Feferman Logic*. In [17] the theory  $\text{FKF}^-$  has been introduced which extends  $\text{WKF}^-$  with a strong implication. We shall define  $\text{FPT}^-$  as its stratified counterpart and  $\text{FPS}^-$  as the respective theory of satisfaction.

**Convention 11.** Let  $\mathcal{L}_{\text{PA}}^{\rightarrow}$  be the language of PA extended with a new primitive symbol  $\rightarrow$ . From now on our metavariables  $\phi, \phi(x), t$  etc. range over the syntactic objects of  $\mathcal{L}_{\text{PA}}^{\rightarrow}$ .

**Remark 190.** Let us note that we have added  $\rightarrow$  to the arithmetised language (the object-language for our truth theories) while it need not be present in the language of theories we work in. In the language we work in, we may treat  $\phi \rightarrow \psi$  as the abbreviation of  $\neg\phi \vee \psi$  (hence in the external language  $\rightarrow$  obeys the law of classical logic.)

**Definition 191** ( $\text{FPT}^-$ ).  $\text{FPT}^-$  consists of all  $\mathcal{L}_{\text{PA}}^{\rightarrow}$ -variants of axioms of  $\text{WPT}^-$  together with the following specific axioms for  $\rightarrow$ :

$$(\rightarrow) \quad \forall\phi\forall\psi \quad (T(\phi \rightarrow \psi) \equiv ((T(\phi) \wedge T(\psi)) \vee T(\neg\phi)))$$

$$\neg(\rightarrow) \quad \forall\phi\forall\psi \quad (T(\neg(\phi \rightarrow \psi)) \equiv (T(\phi) \wedge T(\neg\psi)))$$

**Definition 192** ( $\text{FPS}^-$ ).  $\text{FPS}^-$  consists of all  $\mathcal{L}_{\text{PA}}^{\rightarrow}$ -variants of axioms of  $\text{WPS}^-$  together with the following specific axioms for  $\rightarrow$ :

$$(\rightarrow)_S \quad \forall\phi(\bar{x})\forall\psi(\bar{x})\forall\alpha \in \text{Asn}(\phi, \psi) (S(\phi \rightarrow \psi, \alpha) \equiv ((S(\phi, \alpha \upharpoonright) \wedge S(\psi, \alpha \upharpoonright)) \vee S(\neg\phi, \alpha \upharpoonright)))$$

$$\neg(\rightarrow)_S \quad \forall\phi(\bar{x})\forall\psi(\bar{x})\forall\alpha \in \text{Asn}(\phi, \psi) (S(\neg(\phi \rightarrow \psi), \alpha) \equiv (S(\phi, \alpha \upharpoonright) \wedge S(\neg\psi, \alpha \upharpoonright)))$$

Now we are about to prove the proposition implying that also the pairs

- $\text{FPT}_0$  and  $\text{CT}_0$ ;
- $\text{FPS}_0$  and  $\text{CS}_0$ ;

are mod  $\mathcal{L}_{\text{PA}}$  notational variants. We shall prove it for the case of satisfaction theories, the proof for their truth variants being fully analogous.

**Proposition 193.**  $\text{WPS}_+^-$  is a translational subtheory of  $\text{FPS}^-$ .

*Proof.* The proof is straightforward: working in  $\text{FPS}^-$ , let us define the translation  $*$  for arbitrary formula  $\phi$ , variable  $v$  and term  $t$  by putting:

$$(\{\exists v : v < t\}\phi)^* = \exists v \neg(v < t \rightarrow \neg\phi^*)$$

We also let  $*$  commute with other connectives and being identity on atomic formulae. Translation defined in such a way is obviously  $\mathcal{L}_{\text{PA}}$ -conservative. Now we argue in  $\text{FPS}^-$ : for arbitrary formula  $\phi$ , variable  $v$ , term  $t$  and an assignment  $\alpha \in \text{Asn}(\{\exists v : v < t\}\phi)$  we have (we use the same convention as previously writing  $S_\gamma(\theta)$  for  $S(\theta, \gamma \upharpoonright)$ ):

$$\begin{aligned} S((\{\exists v : v < t\}\phi)^*, \alpha) &\equiv S(\exists v \neg(v < t \rightarrow \neg\phi^*), \alpha) \\ &\equiv \text{tot}_v(\neg(v < t \rightarrow \neg\phi^*), \alpha) \wedge \exists\beta \sim_v \alpha \quad S_\beta(\neg(v < t \rightarrow \neg\phi^*)) \end{aligned}$$

Now we analyse both conjuncts separately. We have:

$$\text{tot}_v(\neg(v < t \rightarrow \neg\phi^*), \alpha) \equiv \forall\beta \sim_v \alpha \left( S_\beta(\neg(v < t \rightarrow \neg\phi^*)) \vee S_\beta(v < t \rightarrow \neg\phi^*) \right)$$

The sentence on the right-hand side is equivalent to

$$\forall\beta \sim_v \alpha \left( (\beta(v) < (t)_\beta^\circ \wedge S_\beta(\phi^*)) \vee ((\beta(v) < (t)_\beta^\circ \wedge S_\beta(\neg\phi^*)) \vee \beta(v) \geq (t)_\beta^\circ) \right)$$

The above is in turn equivalent to

$$\forall\beta \sim_v \alpha \left( S_\beta(v < t) \rightarrow (S_\beta(\phi^*) \vee S_\beta(\neg\phi^*)) \right)$$

and hence to  $\text{tot}_v(v < t, \phi^*, \alpha)$ . Since  $v < t$  is a standard formula, then we also have  $\text{tot}_v(v < t, \alpha)$ . Let us now turn to the second conjunct: we have

$$\exists\beta \sim_v \alpha S_\beta(\neg(v < t \rightarrow \phi^*)) \equiv \exists\beta \sim_v \alpha S_\beta(v < t) \wedge S_\beta(\neg\neg\phi^*)$$

and the last sentence is equivalent to

$$\exists\beta \sim_v \alpha \beta(v) < (t)_\beta^\circ \wedge S_\beta(\phi^*)$$

which ends this part of the proof. Let us occupy with  $\neg\mathbf{G}_b\exists_S$ : we have (for arbitrary formula  $\phi$ , variable  $v$ , term  $t$  and an assignment  $\alpha \in \text{Asn}(\{\exists v : v < t\}\phi)$ )

$$S(\neg(\neg\{\exists v : v < t\}\phi))^*, \alpha) \equiv S(\neg\exists v \neg(v < t \rightarrow \neg\phi^*), \alpha)$$

Now, by the compositional axioms in  $\text{FPS}^-$  we have:

$$S(\neg\exists v \neg(v < t \rightarrow \neg\phi^*), \alpha) \equiv \forall\beta \sim_v \alpha S_\beta(v < t \rightarrow \neg\phi^*)$$

and the last sentence is equivalent to

$$\forall\beta \sim_v \alpha \left( (\beta(v) < (t)_\beta^\circ \wedge S_\beta(\neg\phi^*)) \vee \beta(v) \geq (t)_\beta^\circ \right)$$

This in turn can equivalently be written as

$$\forall\beta \sim_v \alpha (\beta(v) < (t)_\beta^\circ \rightarrow S_\beta(\neg\phi^*))$$

which is the desired translation of  $\neg\mathbf{G}_b\exists_S$ . □

**Corollary 194.**  $\text{FPS}_0$  and  $\text{CS}_0$  are mod  $\mathcal{L}_{\text{PA}}$  notational variants. The same holds for  $\text{FPT}_0$  and  $\text{CT}_0$ .

## 5.2 Disjunctive Correctness and Internal Induction

In this section we shall occupy with apparently the weakest extensions of  $\text{PT}^-$  and  $\text{WPT}^-$  (from those mentioned in Chapter 3): those resulting by adding (appropriate forms of) Disjunctive Correctness and Internal Induction. We shall show that, contrary to what happened previously, the two extensions dramatically differ in strength. We start with a Strong Kleene logic.

## 5.2.1 Strong Kleene Case

Quite surprisingly  $PT^- + DC + INT$  is a very strong theory. Even more surprisingly, the internal logic of this theory is fully classical:

**Theorem 195.**  $PT^- + DC + INT \vdash CT_0$

By "Many Faces" theorem (Theorem 170), it is enough to show that

$$PT^- + DC + INT \vdash CT^-$$

In the following, let us abbreviate  $PT^- + DC + INT$  with simply  $ePT^-$ . We shall show that this theory proves both Tot and Cons axioms, which suffices by Proposition 98, point 3. As in the case of  $PT_0$ , we will establish this by (a kind of) induction on the build-up of formulae. It turns out that in the presence of the axioms for generalised disjunctions the internal induction is a sufficient tool to legitimise this method. As previously, we start with proving two lemmata which show that  $ePT^-$  proves that the truth predicate commutes with blocks of bounded quantifiers. We will need a version of Definition 143 adapted to the present needs:

**Definition 196 (PA).**

1. As previously first lower case letters of Greek alphabet,  $\alpha, \beta, \gamma$ , range over *assignments* i.e. functions mapping variables to numbers. If  $\phi$  is a formula and  $\beta$  an assignment, then we say that  $\beta$  is an *assignment for  $\phi$*  if and only if  $\beta$  assigns values to *all* variables which have free occurrence in  $\phi$  (but it might assigns value to variables which do not occur in  $\phi$ ). The set of assignments for  $\phi$  will be denoted by  $Asn(\phi)$  (note that we are locally changing Definition 87).
2. If  $\phi$  is any formula and  $t$  is any term,  $bucl(\phi, t)$  denotes the *bounded universal closure* of  $\phi$  with respect to  $t$  i.e. the *formula*

$$\forall v_{i_k} < t \forall v_{i_{k-1}} < t \dots \forall v_{i_0} < t(\phi)$$

where  $v_{i_1}, \dots, v_{i_k}$  are all free variables of  $\phi$  listed in the order of decreasing indices (for the sake of determinateness only) and, as usual,

$$\forall v < t\psi$$

is a short for  $\forall v(\neg(v < t) \vee \psi)$  (for arbitrary variable  $v$ , term  $t$  and formula  $\psi$ ). Moreover let  $bucl(x, \phi, t)$  denote the formula resulting from  $\phi$  by bounding last  $x$  free variables (first  $x$  variables with least indices) of  $\phi$  by universal quantifiers bounded by  $t$ .

3. If  $\phi$  is any formula and  $t$  is any term, then  $becl(\phi, t)$  denotes the bounded existential closure of  $\phi$  with respect to  $t$  defined analogously to  $bucl(\phi, t)$ .  $becl(x, \phi, t)$  is defined analogously to  $bucl(x, \phi, t)$ .
4. In contrast to previous sections, it will be convenient to isolate a class of constant assignments together with a specific notation to deal with them. If  $v$  is any variable, then for any  $x$ ,  $\xi_x^v$  denotes the unique assignment whose domain is  $\{v\}$  and which sends  $v$  to  $x$ . If  $\phi$  is any formula, then for every  $x$ ,  $\xi_x^\phi$  denotes the unique assignment whose domain is  $FV(\phi)$  and which maps all the free variables of  $\phi$  to  $x$ .

**Remark 197.** In some contexts, more than one substitution in the formula might be involved and order of their application is important. For example, suppose

$$\begin{aligned}\phi &= \forall v_1 \exists v_0 (v_2 + v_1 = v_0) \\ \psi &= \exists v_0 (v_2 + v_1 = v_0)\end{aligned}$$

and  $\beta$  is an assignment mapping  $v_1$  to 1 and  $v_2$  to 2. Then  $\beta \in \text{Asn}(\phi)$  and

$$\phi[\beta] = \forall v_1 \exists v_0 ((1 + (1 + 0)) + v_1 = v_0)$$

and in  $\text{PT}^-$  we have

$$T(\phi[\beta]) \equiv \forall x T(\psi[\xi_x^{v_1}][\beta])$$

Now for  $x \neq 1$   $\psi[\xi_x^{v_1}][\beta]$  is not equal to  $\psi[\beta][\xi_x^{v_1}]$ , since the latter is equivalent to simply

$$\exists v_0 ((1 + (1 + 0)) + (1 + 0) = v_0).$$

Also note that writing  $[\beta]$  to the right of a formula means that values of  $\beta$  are substituted in the whole formula to the left. In particular, in some cases it might happen that

$$\forall x (\theta)[\beta]$$

is different from

$$\forall x (\theta[\beta])$$

We shall consistently be using parentheses to disambiguate such expressions.

The proof of all the following lemmata is based on a well-known technique of eliminating bounded quantifiers in favour of generalised conjunctions, which is possible in presence of the "disjunctively correct" truth predicate. This technique was used, for example, by Cezary Cieśliński in the proof of implication (5)  $\Rightarrow$  (1) in "Many Faces" Theorem, Ali Enayat in the proof of the implication (6)  $\Rightarrow$  (1) in "Many Faces" Theorem and Bartosz Wcisło in his proof of non-conservativity of  $\text{CT}_0$ . We only adapt it to the present context.

**Lemma 198.** *The following sentence is provable in  $e\text{PT}^-$*

$$\forall \phi (\bar{w}) \forall y (T\text{bucl}(\phi, \underline{y}) \equiv \forall \beta \in \text{Asn}(\phi) (\beta \preceq \xi_{y-1}^\phi \rightarrow T\phi[\beta]))$$

*Proof.* We work in  $e\text{PT}^-$ . Let us fix a formula  $\phi$  and  $a$ . Let  $b$  be the number of free variables of  $\phi$ . Let us define

$$\text{Suff}(w_0, w_1) := ((w_0 \leq \underline{b}) \wedge w_1 = \text{bucl}(w_0, \phi, \underline{a}))$$

( $\text{Suff}(w_0, w_1)$  is a formula in the sense of  $e\text{PT}^-$ ) and any  $x$  such that  $\exists z T(\text{Suff}(z, \underline{x}))$  will be called a *suffix* of  $\phi$  for short (being more precise such  $x$  is a suffix of  $\text{bucl}(\phi, \underline{a})$ .) We shall prove the left-to-right implication first, so let us assume that

$$T\text{bucl}(\phi, \underline{a}) \tag{L-TO-R}$$

holds. Let us fix arbitrary  $\beta \preceq \xi_{a-1}^\phi$ . For  $n \leq b$  let

$$\phi_n := \begin{cases} w_0 = \text{bucl}(0, \phi, \underline{a}) \wedge \phi[\beta], & \text{if } i = 0 \\ w_0 = \text{bucl}(n, \phi, \underline{a}) \wedge \forall v_{i_{n-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)[\beta] & \text{if } n > 0 \end{cases}$$

and, using Definition 129

$$\psi(w_0) = \bigvee_{n < b} \phi_n$$

Hence,  $\psi(w_0)$  can be presented in the following way:

$$\begin{aligned} (w_0 = \underbrace{\forall v_{i_{b-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)} \wedge \forall v_{i_{b-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)) \vee \\ \vee ((w_0 = \underbrace{\forall v_{i_{b-2}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)} \wedge \forall v_{i_{b-2}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)[\beta]) \vee \\ \dots \\ \vee ((w_0 = \underbrace{\forall v_{i_0} < \underline{a}(\phi)} \wedge \forall v_{i_0} < \underline{a}(\phi)[\beta]) \vee \\ \vee (w_0 = \underline{\phi} \wedge \phi[\beta]) \dots) \end{aligned}$$

The following claim exhibits the crucial use of Disjunctive Correctness axiom.

**Claim 3.** For every  $x$  which is a suffix of  $\phi$  we have

$$T(\psi(\underline{x})) \equiv T(x[\beta])$$

*Proof of Claim 3.* Let  $x$  be a suffix of  $\phi$ , i.e. for some  $c$  we have

$$x = \text{bucl}(c, \phi, \underline{a})$$

By Disjunctive Correctness axiom and the definition of  $\psi$  we have

$$T(\psi(\underline{x})) \equiv (\exists n \leq b \ T(\phi_n(\underline{x})))$$

For each  $n$ ,  $T(\phi_n(\underline{x}))$  is by definition equal to

$$T(\underline{x} = \underbrace{\forall v_{i_{n-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)} \wedge \forall v_{i_{n-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)[\beta])$$

or  $T(\underline{x} = \underline{\phi} \wedge \phi[\beta])$  if  $n = 0$ . Hence, by the axiom for conjunction:

$$T(\phi_n(\underline{x})) \equiv (T(\underline{x} = \underbrace{\forall v_{i_{n-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)} \wedge \forall v_{i_{n-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)[\beta]))$$

If  $y \neq c$  then

$$T(\underline{x} = \underbrace{\forall v_{i_{y-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)})$$

is false (by compositional axioms for atomic sentences in  $\text{PT}^-$ ) and, consequently, we must have

$$T(\psi(\underline{x})) \equiv T(\underline{x} = \underbrace{\forall v_{i_{c-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)} \wedge \forall v_{i_{c-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)[\beta])$$

By our choice of  $x$ ,  $T(\underline{x} = \ulcorner \forall v_{i_{c-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi) \urcorner)$  is true and hence we are left with

$$T(\psi(\underline{x})) \equiv T(\forall v_{i_{c-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)[\beta])$$

which by the choice of  $x$  is the same as

$$T(\psi(\underline{x})) \equiv T(x[\beta])$$

□

Now let us define

$$\theta(w_2) := \forall w_0 < w_2 \forall w_1 (\text{Suff}(b - w_0, w_1) \rightarrow \psi(w_1))$$

Let us observe that  $\theta(w_2)$  is a formula of  $\mathcal{L}_{\text{PA}}$ . Intuitively,  $\theta(w_2)$  says that all suffixes of  $\phi$  with more than  $b - w_2$  first quantifiers (equivalently less than  $w_2$  first quantifiers erased) satisfy  $\psi$ . We shall demonstrate that  $\forall y T(\theta(\underline{y}))$  holds which would end our proof. Indeed, suppose we know that  $\forall y T(\theta(\underline{y}))$  holds. Then, in particular, we have  $T(\theta(\underline{b+1}))$  and hence by compositional axioms of  $\text{PT}^-$  and the fact that  $\phi = \text{bucl}(b, \phi)$

$$T(\psi(\phi[\beta])),$$

and  $T(\phi[\beta])$  follows by Claim 3.

To this end, we will use INT to demonstrate  $\forall y T(\theta(\underline{y}))$ . The base step is very easy: we have  $T(\theta(\underline{0}))$ , since  $\theta(0)$  is equal to

$$\ulcorner \forall w_0 (\neg(w_0 < \underline{0}) \vee \forall w_1 (\text{Suff}(w_0, w_1) \rightarrow \psi(w_1))) \urcorner$$

and in  $\text{PT}^-$  the first disjunct is true for every  $x$ . We shall demonstrate that

$$\forall y (T(\theta(\underline{y})) \rightarrow T(\theta(\underline{y+1}))).$$

So let us fix any  $y$  and suppose that  $T(\theta(\underline{y}))$  holds. If  $y > b$ , then for every  $x$  we have

$$T(\neg \text{Suff}(\underline{y}, \underline{x})).$$

Consequently, we have  $T(\forall w_1 (\neg \text{Suff}(\underline{y}, w_1) \vee \psi(w_1)))$ . By our assumption we have also  $\forall z < y T(\forall w_1 (\neg \text{Suff}(\underline{z}, w_1) \vee \psi(w_1)))$  and hence  $T(\theta(\underline{y+1}))$ . So assume  $y \leq b$  and  $T(\theta(\underline{y}))$  holds. We have to demonstrate  $T(\theta(\underline{y+1}))$ , i.e.

$$T\left(\left(\forall w_0 (\neg(w_0 < \underline{y+1}) \vee \forall w_1 (\text{Suff}(w_0, w_1) \rightarrow \psi(w_1)))\right)\right)$$

Let us fix  $x$  of the form

$$\forall v_{i_{(b-z)-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)[\beta]$$

If  $z < y$ , then  $T(\psi(\underline{x}))$  follows by our induction assumption. So let us assume that  $z = y$  (let us stipulate that for  $y = b$  the quantifier prefix of the above formula is empty). By Claim 3 it is enough to demonstrate

$$T(x).$$

If  $y = 0$ , then the above holds by our initial assumption, since  $\text{bucl}(b, \phi, \underline{a})[\beta] = \text{bucl}(\phi, \underline{a}, [])\beta$  and the latter is true (in the sense of  $T$ ) by L-TO-R. If  $y > 0$ , then  $y - 1 < b$  and, since by induction assumption  $T(\theta(\underline{y}))$ , for  $x'$  of the form

$$\forall v_{i_{b-(y-1)-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)[\beta].$$

we have  $T(\psi(\underline{x}'))$ . Once again invoking Claim 3 it follows that

$$T(\forall v_{i_{b-y}} < \underline{a} \forall v_{i_{b-y-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)[\beta])$$

By compositional axioms of  $PT^-$  we have:

$$\forall z < a \ T(\forall v_{b-y-1} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi[\xi_z^{v_{b-y}}]))[\beta]$$

In particular, since  $\beta(v_{b-y}) < a$  then we have

$$T(\forall v_{b-y-1} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi[\xi_{\beta(v_{b-y})}^{v_{b-y}}]))[\beta]$$

Since

$$\forall v_{b-y-1} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi[\xi_{\beta(v_{b-y})}^{v_{b-y}}]))[\beta] = \forall v_{b-y-1} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)[\beta]$$

the proof of one implication is finished.

Let us observe that in the course of induction process, we kept erasing bounded quantifiers one by one and substituting the (numeral naming) the respective value of  $\beta$ . In the proof of the reverse implication of Lemma 198, we will follow the same path, but in the reverse direction. We will use generalised conjunction to simulate bounded universal quantification over assignments.

Let us assume that

$$\forall \beta \in \text{Asn}(\phi) \left( \beta \preceq \xi_{a-1}^\phi \rightarrow T(\phi[\beta]) \right). \quad (\text{L-TO-R})$$

Let  $\beta_0, \dots, \beta_e$  be the increasing enumeration of all assignments for  $\phi$  which are smaller than  $\xi_{a-1}^\phi$  in the sense of  $\preceq$  (by Definition 143 point 5 there can be only finitely many of them). Let us define

$$\gamma_b := \bigwedge_{i \leq e} (\phi)[\beta_i]$$

and for  $c \in [0, b-1]$

$$\gamma_c := \bigwedge_{i \leq e} \forall v_{i_b-c-1} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)[\beta_i]$$

Finally:

$$\psi(w_1) := \bigvee_{c \leq b} ((w_1 = \text{bucl}(c, \phi, \underline{a})) \wedge \gamma_c)$$

Now,  $\psi(x)$  is much harder to depict than its antecedent. The crucial point is that we may use generalised conjunctions instead of bounded quantifier  $\forall \beta \preceq \xi_{a-1}^\phi$  and in such a way present  $\psi(w_2)$  as an arithmetical formula. Now let us make the following

**Claim 4.** If  $x$  is any suffix of  $\phi$ , then

$$T(\psi(\underline{x})) \equiv \forall \beta \preceq \xi_a^\phi (\beta \in \text{Asn}(x) \rightarrow T(x[\beta]))$$

*Proof of Claim 4.* By a reasoning fully analogous to the one used in the proof of Claim 3 we see that if  $x = \text{bucl}(c, \phi, \underline{a})$ , then

$$T(\psi(\underline{x})) \equiv T(\gamma_c)$$

By the truth conditions for generalised conjunction we have:

$$T(\gamma_c) \equiv \forall \beta \preceq \xi_{a-1}^\phi (\beta \in \text{Asn}(\text{bucl}(c, \phi, \underline{a})) \rightarrow T(\text{bucl}(c, \phi, \underline{a})[\beta]))$$

hence by the choice of  $x$  it holds that

$$T(\psi(x)) \equiv \forall \beta \preceq \xi_{a-1}^\phi (\beta \in \text{Asn}(x) \rightarrow T(x[\beta]))$$

□

Let us define:

$$\theta(w_2) := \forall w_0 < w_2 \forall w_1 (\text{Suff}(w_0, w_1) \rightarrow \psi(w_1))$$

Let us observe that once again it is enough to demonstrate that  $\forall y T(\theta(\underline{y}))$  holds. Indeed, assuming this, by *dictum de omni* we get  $T(\theta(\underline{b+1}))$  and, consequently,  $T(\psi(\text{bucl}(\phi, \underline{a})))$  (choosing  $z = b$  and  $x = \text{bucl}(b, \phi, \underline{a}) = \text{bucl}(\phi, \underline{a})$ ). Now by Claim4 we have

$$\forall \beta \preceq \xi_{a-1}^\phi (\beta \in \text{Asn}(\text{bucl}(\phi, \underline{a})) \rightarrow T(\text{bucl}(\phi, \underline{a})[\beta]))$$

and since  $\text{bucl}(\phi, \underline{a})$  does not contain any free variables, the external quantifier is superfluous, and this is precisely what we wanted. To demonstrate  $\forall y T(\theta(\underline{y}))$  we again use internal induction on  $y$ .

$T(\theta(\underline{0}))$  holds for the same trivial reasons as previously. Let us fix  $y$  and assume  $T(\theta(\underline{y}))$ . Once again without loss of generality we may assume that  $y \leq b$ . Let us fix arbitrary  $z < y + 1$ . Similarly to the proof of the first implication we have to show that

$$T(\psi(\text{bucl}(z, \phi, \underline{a})))$$

holds. Now, by Claim 4 the above statement is equivalent to

$$\forall \beta \preceq \xi_{a-1}^\phi (\beta \in \text{Asn}(\text{bucl}(z, \phi, \underline{a})) \rightarrow T(\text{bucl}(z, \phi, \underline{a})[\beta])). \quad (*)$$

Now, if  $z = 0$  then this is our assumption L-TO-R, since  $\text{bucl}(0, \phi, \underline{a}) = \phi$ . So let us assume that  $z > 0$ . Then  $z - 1 < y$  and by our induction assumption and Claim 4 we have

$$\forall \beta \preceq \xi_{a-1}^\phi (\beta \in \text{Asn}(\text{bucl}(z-1, \phi, \underline{a})) \rightarrow T(\text{bucl}(z-1, \phi, \underline{a})[\beta])) \quad (**)$$

Let us observe that in order to demonstrate \* it is enough to focus on  $\beta$  assigning values *only* to variables

$$v_{i_{b-1}}, \dots, v_{i_{b-z}}$$

since these are the only free variables of  $\text{bucl}(z, \phi, \underline{a})$ . Let  $\beta'$  be arbitrary such assignment. Then

$$\text{bucl}(z-1, \phi, \underline{a})[\beta']$$

contains precisely one free variable  $v_{i_{b-z}}$ , since this is the variable with least index which is not bounded in  $\text{bucl}(z-1, \phi, \underline{a})$ . But for arbitrary  $d < a$ , by \*\* we have

$$T(\text{bucl}(z-1, \phi, \underline{a})[\beta'][\xi_d^{v_{i_{b-z}}}]$$

hence

$$\forall x < a T(\text{bucl}(z-1, \phi, \underline{a})[\beta'](\underline{x}))$$

In particular, invoking standard compositional axioms in  $\text{PT}^-$  we have

$$T(\text{bucl}(z, \phi, \underline{a})[\beta'])$$

which ends the whole proof. □

The following corollary is an easy consequence

**Corollary 199.** *The following sentence is provable in  $ePT^-$*

$$\forall \phi(\bar{w}) \forall v \notin \text{FV}(\phi) \left( T^\top \forall v \text{bucl}(\phi, v)^\top \equiv (\forall \beta \in \text{Asn}(\phi) T\phi[\underline{\beta}]) \right)$$

The proof of the next lemma follows the same pattern as the proof of Lemma 198, so we allow ourselves some sloppiness. The idea is the same: we use Disjunctive Correctness and Internal Induction to perform induction on the build-up of formulae.

**Lemma 200.**  $ePT^- \vdash \text{Tot}$

*Proof.* We work in  $ePT^-$ . Let us fix arbitrary sentence  $\phi \in \mathcal{L}_{\text{PA}}$  and let  $\rho_0, \dots, \rho_a$  be any enumeration of its subformulae. Let  $v$  be any variable which does not occur in  $\phi$  (either as free or a bounded one). Let us define

$$\begin{aligned} \phi'_i &:= \forall v \text{bucl}(\rho_i \vee \neg \rho_i, v), \text{ for } i \leq a \\ \phi_i(w) &:= w = \underline{\rho_i} \wedge \phi'_i \\ \psi(w) &:= \bigvee_{i \leq a} \phi_i(w) \end{aligned}$$

As in the previous lemmata let us isolate the following:

**Claim 5.** For every subformula  $\xi$  of  $\phi$  we have

$$T(\psi(\underline{\xi})) \equiv \forall \beta \in \text{Asn}(\xi) (T(\xi[\beta]) \vee T(\neg \xi[\beta]))$$

*Proof of Claim 5.* Let us fix  $\xi = \rho_i$  for some  $i$ . As in the proof of Claim 3 we have

$$T(\psi(\underline{\rho_i})) \equiv T(\forall v \text{bucl}(\rho_i \vee \neg \rho_i, v)).$$

Hence, Corollary 199 gives us

$$T(\psi(\underline{\rho_i})) \equiv \forall \beta \in \text{Asn}(\rho_i) (T(\rho_i[\beta]) \vee T(\neg \rho_i[\beta]))$$

which ends the proof of our claim. □

By the above claim it is sufficient to demonstrate that

$$T(\psi(\underline{\phi}))$$

holds. Let us put

$$\theta(w_2) := \forall w_0 < w_2 \forall w_1 \left( (\text{Subf}(\phi, w_1) \wedge \text{Compl}(w_1) = w_0) \rightarrow \psi(w_1) \right)$$

Obviously  $\theta$  is an arithmetical formula. Let  $b$  be the complexity of  $\phi$ . As in the proof of preceding lemmata,  $T(\theta(\underline{b+1}))$  implies  $T(\psi(\underline{\phi}))$ . By internal induction on  $y$  we shall show

$$\forall y T(\theta(\underline{y}))$$

which will end the proof. As usual verifying that  $T(\theta(0))$  is trivial. Let us fix  $y$  and assume  $T(\theta(\underline{y}))$ . In order to show  $T(\theta(\underline{y+1}))$  let us also pick  $z < y+1$  and  $\xi$  such that  $\xi$  is any subformula of  $\phi$  of complexity  $z$ . Let  $i$  be such that  $\xi = \rho_i$ . By Claim 5 it is enough to demonstrate

$$\forall \beta \in \text{Asn}(\rho_i) (T(\rho_i[\beta]) \vee T(\neg \rho_i[\beta])) \quad (*)$$

Let us fix any  $\beta \in \text{Asn}(\rho_i)$ . Clearly we can assume that  $\beta$  assigns values only to the free variables of  $\rho_i$ . Now the proof proceeds by cases: if  $\rho_i$  is atomic, then (\*) follows by compositional axioms for atomic sentences. So let us suppose that the complexity of  $\rho_i$  is non-zero. In such a situation we further distinguish cases depending on the main connective of  $\rho_i$ . Without loss of generality we may assume that  $\rho_i$  does not start with the negation sign. We shall show the steps for  $\vee$  and  $\exists$ . Let  $\rho_i = \rho_k \vee \rho_j$  for some  $k, j \leq a$ . Consequently,  $\max\{\text{Compl}(\rho_k), \text{Compl}(\rho_j)\} < y$ , hence by our induction assumption we obtain

$$T(\psi(\underline{\rho_j})), T(\psi(\underline{\rho_k}))$$

and by Claim 5

$$\forall \beta \in \text{Asn}(\rho_j) (T(\rho_j[\beta]) \vee T(\neg \rho_j[\beta])) \quad (5.2)$$

$$\forall \beta \in \text{Asn}(\rho_k) (T(\rho_k[\beta]) \vee T(\neg \rho_k[\beta])) \quad (5.3)$$

By the compositional axiom for disjunction in  $\text{PT}^-$  (\*) for the fixed  $\beta$  is equivalent to

$$(T(\rho_j[\beta]) \vee T(\rho_k[\beta])) \vee (T(\neg \rho_j[\beta]) \wedge T(\neg \rho_k[\beta]))$$

which clearly follows from (5.2) and (5.3). So let us assume that  $\rho_i = \exists v_j \rho_k$ . In such a case our induction assumption gives us

$$\forall \beta \in \text{Asn}(\rho_k) (T(\rho_k[\beta]) \vee T(\neg \rho_k[\beta])) \quad (5.4)$$

By compositional axioms in  $\text{PT}^-$ , (\*) for the chosen  $\beta$  (we assumed that  $\beta$  assigns values only to the free variables of  $\rho_i$ , hence  $v_j \notin \text{dom}(\beta)$ ), is equivalent to

$$\forall x T(\neg \rho_k[\beta](\underline{x})) \vee \exists x T(\rho_k[\beta](\underline{x})) \quad (5.5)$$

(both above expressions are meaningful since  $\rho_k[\beta]$  contains at most one free variable.) Hence, 5.4 is equivalent to

$$\forall \beta \in \text{Asn}(\rho_i) \forall x \left( T(\rho_k[\beta](\underline{x})) \vee T(\neg \rho_k[\beta](\underline{x})) \right)$$

from which it follows that

$$\forall x \left( T(\rho_k[\beta](\underline{x})) \vee T(\neg \rho_k[\beta](\underline{x})) \right)$$

where  $\beta$  is our fixed valuation. The above clearly implies (5.5) and our proof is finished.  $\square$

Proof of the lemma below follows exactly the dual pattern to proof of Lemma 198: instead of bucl we use becl.

**Lemma 201.** *The following sentence is provable in  $e\text{PT}^-$*

$$\forall \phi(\bar{w}) \forall y \ (T\text{becl}(\phi, \underline{y}) \equiv \exists \beta \in \text{Asn}(\phi) (\beta \preceq \xi_{y-1}^\phi \wedge T\phi[\beta]))$$

*Sketch of the proof.* We fix  $\phi$ ,  $a$  and define  $b$  to be the number of free variables of  $\phi$ . Let

$$\text{Suff}(w_0, w_1) := ((w_0 \leq \underline{b}) \wedge w_1 = \text{becl}(w_0, \phi, \underline{a})).$$

We show the right-to-left implication first. Let us fix arbitrary  $\beta \in \text{Asn}(\phi)$  such that  $\beta \preceq \xi_{a-1}^\phi$  and  $T(\phi[\beta])$  and define

$$\begin{aligned} \phi_n(w_1) &:= (w_1 = \text{becl}(n, \phi, \underline{a}) \wedge \exists v_{i_{n-1}} < \underline{a} \dots \exists v_{i_0} < \underline{a}(\phi)[\beta]) \\ \psi(w_1) &:= \bigvee_{n \leq b} \phi_n(w_1) \end{aligned}$$

Moreover put

$$\theta(w_2) := \forall w_0 < w_2 \forall w_1 (\text{Suff}(w_0, w_1) \rightarrow \psi(w_1))$$

and via argument analogous to the one used in the proof of Lemma 198 show that  $\forall y T(\theta(\underline{y}))$ . To show the converse implication assume that  $T(\text{becl}(\phi, \underline{a}))$  holds. Let  $\beta_0, \dots, \beta_e$  be the enumeration of all elements of  $\text{Asn}(\phi)$  smaller than  $\xi_{a-1}^\phi$  in the sense of  $\preceq$ . For  $c \in [0, b]$  define:

$$\begin{aligned} \gamma_c &:= \bigvee_{i \leq e} \exists v_{i_{b-c-1}} < \underline{a} \dots \forall v_{i_0} < \underline{a}(\phi)[\beta_i] \text{ if } c < b \\ \gamma_c &:= \bigvee_{i \leq e} \phi[\beta_i] \text{ if } c = b \\ \psi(w_1) &:= \bigvee_{c \leq b} ((w_1 = \text{becl}(c, \phi, \underline{a})) \wedge \gamma_c) \end{aligned}$$

Define  $\theta(w_2) = \forall w_0 < w_2 \forall w_1 (\text{Suff}(\underline{b} - w_0, w_1) \rightarrow \psi(w_1))$  and using internal induction show that  $\forall y T(\theta(\underline{y}))$ .  $\square$

**Corollary 202.** *The following sentence is provable in  $e\text{PT}^-$*

$$\forall \phi(\bar{w}) \forall v \notin \text{FV}(\phi) \left( T^\Gamma \exists v \text{becl}(\phi, v)^\Gamma \equiv (\exists \beta \in \text{Asn}(\phi) T\phi[\underline{\beta}]) \right)$$

**Lemma 203.**  $e\text{PT}^- \vdash \text{Cons}$ .

*Proof.* Working in  $e\text{PT}^-$  let us assume that for some sentence  $\phi$  we have

$$T(\phi \wedge \neg \phi)$$

Let  $\rho_0, \dots, \rho_a$  be any enumeration of subformulae of  $\phi$  and let us suppose that  $b = \text{Compl}(\phi)$ . Define

$$\begin{aligned} \phi'_i &:= \exists v \text{becl}(\rho_i \wedge \neg \rho_i) \\ \phi_i(x) &:= (x = \rho_i \wedge \phi'_i) \\ \psi(x) &:= \bigvee_{i \leq a} \phi_i(x) \end{aligned}$$

Exactly as in the proof of Claim 5 we can demonstrate the following:

**Claim 6.** For every subformula  $\xi$  of  $\phi$  we have

$$T(\psi(\underline{\xi})) \equiv \exists \beta \in \text{Asn}(\xi) (T(\xi[\beta]) \wedge T(\neg\xi[\beta]))$$

Let us now put

$$\theta(w_2) := \exists w_0 \leq \underline{b} - w_2 \exists w_1 (\text{Subf}(\phi, w_1) \wedge \text{Compl}(w_1) = w_0 \wedge \psi(w_1))$$

Let us recall that by our convention, if  $a < z$ , then  $a - z = 0$ . Intuitively formula  $\theta(\underline{y})$  says that below the  $y$ -th level of the syntactic tree of  $\phi$  we can see a subformula of  $\phi$  which is inconsistent. Let us observe that  $T(\theta(\underline{b}))$  implies that there exists an inconsistent formula of complexity 0, which is impossible, since such formulae are atomic and  $\text{PT}^-$  proves that atomic formulae are consistent. We will demonstrate that  $\forall y T(\theta(\underline{y}))$  by internal induction on  $y$ . If  $y = 0$  then  $T(\theta(\underline{0}))$  is true by our assumption that  $T(\phi \wedge \neg\phi)$ . Let us fix  $y$  and assume  $T(\theta(\underline{y}))$  holds. In particular, there exists a subformula  $\xi$  of  $\phi$  of complexity at most  $b - y$  such that

$$T(\psi(\underline{\xi}))$$

holds. By Claim 6 we have

$$\exists \beta \in \text{Asn}(\xi) (T(\xi[\beta]) \wedge T(\neg\xi[\beta])) \quad (5.6)$$

Let us fix any  $\beta$  witnessing the above existential statement. Without loss of generality let us assume that  $\beta$  is defined only on the variables which have free occurrence in  $\xi$ . As we already observed,  $\xi$  cannot be atomic, since such formulae are consistent. Hence,  $\xi$  is compound and we distinguish cases depending on what is the main logical symbol in  $\xi$ . Without loss of generality we may assume that  $\xi$  does not start with the negation sign, for if  $\xi = \neg\xi'$ , then  $T(\neg\xi'[\beta]) \wedge T(\neg\neg\xi'[\beta])$  is equivalent to

$$T(\neg\xi'[\beta]) \wedge T(\xi'[\beta])$$

and  $\xi'$  is of complexity at most  $b - y - 1$ . We shall show the steps for  $\vee$  and  $\exists$ . Let us assume first that for some  $i, j \leq a$   $\xi = \rho_i \vee \rho_j$ . By (5.6) and the axioms for the negation in  $\text{PT}^-$  we get:

$$(T(\rho_i[\beta]) \vee T(\rho_j[\beta])) \wedge (T(\neg\rho_i[\beta]) \wedge T(\neg\rho_j[\beta])) \quad (5.7)$$

The above clearly implies

$$(T(\rho_i[\beta]) \wedge T(\neg\rho_i[\beta])) \vee (T(\rho_j[\beta]) \wedge T(\neg\rho_j[\beta]))$$

Without loss of generality assume that we have  $T(T(\rho_i[\beta]) \wedge T(\neg\rho_i[\beta]))$ . Once again by Claim 6 we obtain

$$T(\psi(\underline{\rho_i}))$$

Since  $\rho_i$  is of complexity smaller than  $\xi$ , then it is of complexity at most  $a - (y + 1)$  and in particular, we have

$$T(\theta(\underline{y + 1}))$$

which ends the step for  $\vee$ .

Let us now suppose that  $\xi = \exists v_j \rho_i$  for some  $i \leq a$  and some  $j$ . By axioms for  $\exists$  in  $\text{PT}^-$  we have

$$\exists x T(\rho_i[\beta](x)) \wedge \forall x T(\neg \rho_i[\beta](x))$$

In particular, for some  $d$  we have

$$T(\rho_i[\beta](d)) \wedge T(\neg \rho_i[\beta](d))$$

Let us put  $\gamma = \beta \cup \{\langle v_j, d \rangle\}$ . Then  $\gamma \in \text{Asn}(\rho_i)$  and we get:

$$T(\rho_i[\gamma]) \wedge T(\neg \rho_i[\gamma])$$

Hence, by Claim 6:  $T(\psi(\rho_i))$ . Since  $\rho_i$  is of complexity smaller than  $\xi$ , then it is of complexity at most  $a - (y + 1)$  and in particular, we have

$$T(\theta(y + 1))$$

which ends the step for  $\exists$  and the whole proof.  $\square$

The above lemmata, together with Proposition 98, point 3, complete the proof of Theorem 204.

### Digression 1: Restricting Internal Induction

We shall show that the full strength of internal induction is indeed needed to cross the Tarski Boundary. The proof will be an easy modification of the one from [5] which demonstrated Lemma 117.

**Theorem 204.**  $\text{PT}^- + \text{DC} + \text{INT}(\text{tot})$  is proof-theoretically conservative over PA.

We shall modify operator  $\Theta$  from Definition 115 adding the conditions for generalised disjunction. Let  $\mathcal{M} \models \text{PA}$ .

$$\begin{aligned} \Theta_{\text{DC}}^{\mathcal{M}}(\phi, A) := & \mathcal{M} \models \exists s, t [\phi = (s = t) \wedge s^\circ = t^\circ] \\ & \vee \mathcal{M} \models \exists s, t [\phi = \neg(s = t) \wedge s^\circ \neq t^\circ] \\ & \vee \exists \psi \in \text{Sent}_{\mathcal{M}} [\mathcal{M} \models \phi = \neg\neg\psi] \wedge \psi \in A \\ & \vee \exists \psi_1, \psi_2 \in \text{Sent}_{\mathcal{M}} [\mathcal{M} \models \phi = (\psi_1 \vee \psi_2)] \wedge (\psi_1 \in A) \vee (\psi_2 \in A) \\ & \vee \exists \psi_1, \psi_2 \in \text{Sent}_{\mathcal{M}} [\mathcal{M} \models \phi = \neg(\psi_1 \vee \psi_2)] \wedge (\neg\psi_1 \in A) \wedge (\neg\psi_2 \in A) \\ & \vee \exists c \in \text{SetSent}_{\mathcal{M}} [\mathcal{M} \models \phi = \bigvee_{\psi \in c} \psi] \wedge \exists \psi \in c \ \psi \in A \end{aligned} \tag{5.8}$$

$$\vee \exists c \in \text{SetSent}_{\mathcal{M}} [\mathcal{M} \models \phi = \neg \bigvee_{\psi \in c} \psi] \wedge \forall \psi \in c \ \neg\psi \in A \tag{5.9}$$

$$\vee \exists \psi \in \text{Form}_{\mathcal{M}}^1 [\mathcal{M} \models \phi = \exists x \psi] \wedge \exists s \in \text{Tm}^c \ (\psi(s) \in A)$$

$$\vee \exists \psi \in \text{Form}_{\mathcal{M}}^1 [\mathcal{M} \models \phi = \neg \exists x \psi] \wedge \forall s \in \text{Tm}^c \ (\neg\psi(s) \in A)$$

where the only added conditions are (4.8) and (4.9).  $\Gamma_{\text{DC}}^{\mathcal{M}}, \Gamma_{\text{DC}}^{\mathcal{M}}(\alpha), \alpha_{\text{DC}}^{\mathcal{M}}$  are defined as in Definition 115 but using  $\Theta_{\text{DC}}^{\mathcal{M}}$  instead of  $\Theta^{\mathcal{M}}$ . Now, similarly to the unmodified operator  $\Theta$  it is easy to observe that if  $A \subset M$  is any fixpoint of  $\Gamma_{\text{DC}}^{\mathcal{M}}$  then

$$(\mathcal{M}, A) \models \text{PT}^- + \text{DC}.$$

Now we prove a variant of lemma 117 for  $\Theta_{\text{DC}}^{\mathcal{M}}$ .

**Lemma 205.** *If  $\mathcal{M}$  is recursively saturated, then  $\alpha_{\text{DC}}^{\mathcal{M}} = \omega$ .*

*Proof.* We check that  $\Gamma_{\text{DC}}(\omega)$  is a fixpoint of  $\Gamma$  (and skip the reference to  $\mathcal{M}$ ). All the steps are as in the proof of Lemma 3.9 in [5] except for the generalised disjunctions. The only problematic step is to demonstrate that for every  $c \in \text{SetSent}$  if

$$\forall \phi \in c \ (\neg \phi \in \Gamma_{\text{DC}}(\omega)) \tag{5.10}$$

then  $\neg \bigvee_{\phi \in c} \phi \in \Gamma_{\text{DC}}(\omega)$ . This will certainly be the case if there exists  $n \in \mathbb{N}$  such that

$$\forall \phi \ (\neg \phi \in \Gamma_{\text{DC}}(n)) \tag{5.11}$$

So assume there exists  $c$  such that (5.10) holds, but (5.11) is not the case for every  $n$ . As in the proof of Lemma 3.9 from [5] for each  $n$  there exists an arithmetical formula  $\Gamma_{\text{DC}}^n(x)$  which defines  $\Gamma_{\text{DC}}(n)$ . Let us consider the following set of formulae with parameter  $c$

$$p(x) := \{x \in c \wedge \neg \Gamma_{\text{DC}}^n(\neg x) \mid n \in \omega\}$$

$p(x)$  is clearly recursive. Also, by our assumption it is finitely satisfiable, hence by the choice of  $\mathcal{M}$  there exists  $a \in M$  such that  $a \in c$  and for every  $n$

$$\neg a \notin \Gamma_{\text{DC}}(n).$$

In particular,  $\neg a \notin \Gamma_{\text{DC}}(\omega)$  which contradicts (5.10).  $\square$

A variant of this lemma, for operator  $\Theta^{\mathcal{M}}$ , was proven in [5] (Lemma 3.10). What we give below is just a modification of this proof.

**Lemma 206.** *If  $\alpha_{\text{DC}}^{\mathcal{M}} = \omega$ , then  $(\mathcal{M}, \Gamma_{\text{DC}}(\omega)) \models \text{PT}^- + \text{DC} + \text{INT}(\text{tot})$ .*

*Proof.* Let us fix  $\mathcal{M}$  such that  $\alpha_{\text{DC}}^{\mathcal{M}} = \omega$ . We already know that

$$(\mathcal{M}, \Gamma_{\text{DC}}(\omega)) \models \text{PT}^- + \text{DC}$$

so we have to verify the internal induction for total formulae. The proof is the same as in Lemma 3.10 in [5], but we reprove it for Reader's convenience. Let us fix arbitrary  $\phi(v) \in \text{Form}_{\mathcal{M}}^{\leq 1}$  and assume that it is total i.e.

$$\forall x (T(\phi(\underline{x})) \vee T(\neg \phi(\underline{x})))$$

holds in  $(\mathcal{M}, \Gamma_{\text{DC}}(\omega))$ . Since  $\Gamma_{\text{DC}}(\omega)$  is a fixpoint of  $\Gamma_{\text{DC}}$ , then we have

$$(\forall x (\phi(\underline{x}) \vee \neg \phi(\underline{x}))) \in \Gamma_{\text{DC}}(n)$$

for some  $n \in \omega$ . Let  $k$  be the least  $n$  for which the above holds. In particular, we have

$$\forall x (\phi(\underline{x}) \in \Gamma_{\text{DC}}(k-1) \vee \neg\phi(\underline{x}) \in \Gamma_{\text{DC}}(k-1))$$

( $k$  cannot equal 0, since the above is not an atomic formula). Since  $(\mathcal{M}, \Gamma_{\text{DC}}(\omega)) \models \text{Cons}$ , then we in fact have

$$\forall x \left( (\phi(\underline{x}) \in \Gamma_{\text{DC}}(k-1)) \equiv (\phi(\underline{x}) \in \Gamma_{\text{DC}}(\omega)) \right)$$

Since  $\Gamma_{\text{DC}}(k-1)$  is definable in  $\mathcal{M}$  and for the formula defining  $\Gamma_{\text{DC}}(k-1)$  in  $\mathcal{M}$  satisfies the instantiation of the induction axiom, then the above observation ends our proof.  $\square$

The above lemma completes the proof of Theorem 204 (we invoke Proposition 59). Let us observe that one consequence of the above theorem is that  $\text{PT}^- + \text{DC}$  is proof-theoretically conservative over PA. Since  $\text{PT}^- + \text{INT}$  is a subtheory of  $\text{CT}^- + \text{INT}$ , then by Theorem 140, it is also proof-theoretically conservative. Hence, we see two truth principles DC and INT, each of which when added separately to  $\text{PT}^-$  generate a proof-theoretically conservative theory but adding them *together* results in a very strong theory. Consequently, it might be argued that this is really the interplay between the two principles that is needed to obtain the strengthening (in the proof-theoretical sense) of  $\text{PT}^-$ .

### 5.2.2 Weak Kleene Case

We shall now occupy with the strength of  $\text{WPT}^- + \text{DC} + \text{INT}$ . We shall slightly strengthen both principles  $G\bigvee_{wk}$  and  $G\neg\bigvee$ , which, since we are making a conservativity claim, will also (also slightly) strengthen the obtained result. Let us introduce the following definition:

**Definition 207** (PA). Let

$$x = \bigvee^* y$$

be an arithmetical formula (of two variables  $x$  and  $y$ ) strongly representing the relation:  $y$  is a set of sentences and  $x$  is a disjunction of formulae from  $y$  *parenthesized in some way*.

Let us note that for a fixed set of sentences  $y$ , there might be no unique  $x$  such that

$$x = \bigvee^* y$$

For example, working in PA, if for some  $\psi_0, \psi_1, \psi_2, y = \{\psi_0, \psi_1, \psi_2\}$ , then we have

$$\begin{aligned} (\psi_0 \vee (\psi_1 \vee \psi_2)) &= \bigvee^* y \\ ((\psi_0 \vee \psi_1) \vee \psi_2) &= \bigvee^* y \end{aligned}$$

Now we introduce stronger version of the two above correctness principles. The only novelty is that the resulting theory will prove that the truth of a (generalised) disjunction does not depend on the chosen way of parenthesizing it.

$$\forall x \forall y \left( (x = \bigvee^* y) \rightarrow T(x) \equiv ((\forall \phi \in y \text{ tot}(\phi)) \wedge (\exists \phi \in y T(\phi))) \right) \quad (\text{GV}_{wk}^*)$$

$$\forall x \forall y \left( (x = \bigvee^* y) \rightarrow T(\neg x) \equiv \forall \phi \in y T(\neg \phi) \right) \quad (\text{G}\neg\text{V}_{wk}^*)$$

**Definition 208.**  $\text{WPT}^- + \text{DC} + \text{INT}$  is the theory  $\text{WPT}^- + \text{INT}$  with  $\text{wG}\bigvee^*$  and  $\text{wG}\neg\bigvee^*$  added.

**Theorem 209.**  $\text{WPT}^- + \text{DC} + \text{INT}$  is model-theoretically conservative over PA.

Let us fix arbitrary model  $\mathcal{M}$ . We shall modify operator  $\Theta_{\text{DC}}^{\mathcal{M}}$  and adapt it to the present context. For the sake of readability, we will not introduce new symbol for this new version. For  $A \subseteq M$  we introduce the following abbreviations:

1.  $\text{tot}_v(\phi, A)$  stands for  $\forall x \left( \phi(\underline{x}/v) \in A \vee \neg \phi(\underline{x}/v) \in A \right)$ . As previously if  $\phi$  is a sentence then  $\text{tot}_v(\phi, A)$  is equivalent to simply  $\phi \in A \vee \neg \phi \in A$ .
2.  $\bigvee_{wk}(\phi, \psi, A)$  stands for  $\text{tot}_v(\phi, A) \wedge \text{tot}_v(\psi, A) \wedge (\phi \in A \vee \psi \in A)$ .

We define  $\Theta$  :

$$\begin{aligned} \Theta_{\text{DC}}^{\mathcal{M}}(\phi, A) := & \mathcal{M} \models \exists s, t [\phi = (s = t) \wedge s^\circ = t^\circ] \\ & \vee \mathcal{M} \models \exists s, t [\phi = \neg(s = t) \wedge s^\circ \neq t^\circ] \\ & \vee \exists \psi \in \text{Sent}_{\mathcal{M}} [\mathcal{M} \models \phi = \neg\neg\psi] \wedge \psi \in A \\ & \vee \exists c \in \text{SetSent}_{\mathcal{M}} [\mathcal{M} \models \phi = \bigvee^* c] \wedge (\forall \psi \in c \text{ tot}_v(\psi, A)) \wedge \exists \psi \in c (\psi \in A) \\ & \vee \exists c \in \text{SetSent}_{\mathcal{M}} [\mathcal{M} \models \phi = \neg \bigvee^* c] \wedge \forall \psi \in c (\neg \psi \in A) \\ & \vee \exists \psi \in \text{Form}_{\mathcal{M}}^1 [\mathcal{M} \models \phi = \exists v \psi] \wedge \text{tot}_v(\psi, A) \wedge \exists s \in \text{Tm}^c (\psi(s) \in A) \\ & \vee \exists \psi \in \text{Form}_{\mathcal{M}}^1 [\mathcal{M} \models \phi = \neg \exists v \psi] \wedge \forall s \in \text{Tm}^c (\neg \psi(s) \in A) \end{aligned}$$

Let us observe that we do not need to add a special clause for  $\bigvee$ , since  $\theta \vee \gamma$  can be written simply as

$$\bigvee_{\phi \in c} \phi$$

where  $c = \{\theta, \gamma\}$  (forgetting about the order of formulae occurring in the disjunction).

Let  $\Gamma_{\text{DC}}^{\mathcal{M}}, \Gamma_{\text{DC}}^{\mathcal{M}}(\alpha)$  be as defined in Definition 115 but with the above  $\Theta_{\text{DC}}^{\mathcal{M}}$  instead of  $\Theta^{\mathcal{M}}$ . As in the previous cases, we have the following proposition:

**Proposition 210.** *If  $A$  is any fixpoint of  $\Gamma_{\text{DC}}^{\mathcal{M}}$ , then  $(\mathcal{M}, A) \models \text{WPT}^- + \text{DC}$ .*

Contrary to other operators, the fixpoint of  $\Theta_{\text{DC}}^{\mathcal{M}}$  is always reached after as few steps as possible. We have the following:

**Proposition 211.** *In every  $\mathcal{M} \models \text{PA}$ ,  $\Gamma_{\text{DC}}^{\mathcal{M}}(\omega)$  is a fixpoint of  $\Gamma_{\text{DC}}^{\mathcal{M}}$ .*

In the proof of the above proposition, we shall isolate one lemma that brings to the light the crucial difference between operators  $\Theta$  for Weak and Strong Kleene Logic. Let us redefine the measure  $\text{Compl}$ , defined in Definition 24. We would like to make it possible to count generalised disjunction as, in a sense, one symbol. It would be most convenient to use a modified version of the function  $\text{Compl}(\phi) = n$ . Since it was based on Definition 20, we show how to modify it:

**Definition 212** (Generalized Syntactic Tree; PA). The generalised reduced syntactic tree of a formula  $\phi \in \mathcal{L}_{\text{PA}}$  is a pair  $\langle A^\phi, l^\phi \rangle$  where  $A^\phi$  is a set of (arbitrary) sequences closed under prefixes and  $l^\phi : A^\phi \rightarrow \text{Form}_{\mathcal{L}_{\text{PA}}}$  such that conditions 1', 2 and 4 from Definition 20 are satisfied together with the generalised variant of condition 2:

$$2' \text{ if } \phi = \bigvee_{i < x+1} \psi_i \text{ for some formulae } \psi_0, \psi_1, \dots, \psi_x \text{ then } A^\phi = \bigcup_{i < x+1} i \frown A^{\psi_i}, l^\phi(\varepsilon) = \phi \\ \text{and for every } \sigma \in A^{\psi_i}, l^\phi(i \frown \sigma) = l^{\psi_i}(\sigma).$$

We shall denote the above measure with  $\text{Compl}^*(x)$ .

**Definition 213.** Let  $\mathcal{M}$  be any model of PA and  $\phi \in \text{Form}_{\mathcal{M}}$ . We shall say that a set  $A \subset \mathcal{M}$  *decides*  $\phi$  if for every  $\beta \in \text{Asn}(\phi)$  we have either  $\phi[\beta] \in A$  or  $\neg\phi[\beta] \in A$ .

Now we can prove the following lemma which distinguishes Weak Kleene operator  $\Theta$  from its Strong version:

**Lemma 214.** *Let  $\mathcal{M} \models \text{PA}$  and  $\phi \in \text{Form}_{\mathcal{M}}$ . For every  $n$  we have*

$$\Gamma_{\text{DC}}^{\mathcal{M}}(n) \text{ decides } \phi \iff \text{Compl}^*(\phi) \leq n$$

*Proof.* Let us fix  $\mathcal{M}$  and  $\phi \in \text{Form}_{\mathcal{M}}$  and omit both the superscript  $\mathcal{M}$  and the subscript DC in  $\Gamma_{\text{DC}}^{\mathcal{M}}(n)$ . Let us observe that for every  $n$  both  $\Gamma(n)$  and  $\text{Compl}^*(x) \leq n$  are definable in  $\mathcal{M}$ . We will also use

$$\text{tot}(x, \Gamma(n))$$

(and  $(\text{tot}_v(x, \Gamma(n)))$ ) as an arithmetical formula defining the set of sentences (formulae) which are decided by  $\Gamma(n)$ . We use (external) induction on  $n$  and from now on work in  $\mathcal{M}$ . For  $n = 0$  the thesis follows immediately, since  $\Gamma(0)$  contains exactly (codes of) true atomic sentences and (codes of) negations of false atomic sentences. Let us assume that our thesis holds for  $k$ . Let us fix arbitrary  $\phi \in \text{Form}_{\mathcal{L}_{\text{PA}}}$  and assume first that  $\Gamma(k+1)$  decides  $\phi$ . If  $\phi$  is atomic or negated atomic then it is certainly of complexity at most  $k+1$ . If not, then by considering each connective separately it is easy to demonstrate that  $\Gamma(k)$  decides all immediate subformulae of  $\phi$ . Indeed let us show the  $\bigvee$  case: working in  $\mathcal{M}$  suppose for some set  $c$ ,  $\phi = \bigvee^* c$ . Let us fix arbitrary  $\beta \in \text{Asn}(\phi)$ . By our assumption we have

$$\phi[\beta] \in \Gamma(k+1) \vee \neg\phi[\beta] \in \Gamma(k+1)$$

Without loss of generality let us assume that the first holds. Then by the definition of  $\Gamma(k)$  for every  $\theta \in c$  we have

$$\text{tot}(\theta[\beta], \Gamma(k))$$

( $\beta$  is also an assignment for  $\theta$  by Definition 196) which means that  $\Gamma(k)$  decides  $\theta[\beta]$ , hence by our induction assumption every  $\theta \in c$  is of complexity at most  $k$ . Hence,  $\phi$  is of complexity at most  $k + 1$ .

Let us now show  $\exists$  case. Suppose  $\phi = \exists v\psi$ . Let us fix arbitrary  $\beta \in \text{Asn}(\psi)$ . Let us call  $\beta'$  the restriction of  $\beta$  to the free variables of  $\phi$ . Then by our assumption we have

$$\phi[\beta'] \in \Gamma(k+1) \vee \neg\phi[\beta'] \in \Gamma(k+1)$$

Let us assume that  $\phi[\beta'] \in \Gamma(k+1)$ . In particular, we have

$$\text{tot}_v(\psi[\beta'], \Gamma(k))$$

which implies that  $\Gamma(k)$  decides  $\psi[\beta]$ . In the second case, i.e.  $\neg\phi[\beta'] \in \Gamma(k+1)$ , we have

$$\forall x(\neg\psi[\beta'](\underline{x})) \in \Gamma(k).$$

In particular, since  $\neg\psi[\beta'](\underline{\beta(v)}) = \neg\psi[\beta]$  we have  $\neg\psi[\beta] \in \Gamma(k)$ . Hence,  $\Gamma(k)$  decides  $\psi$  and by our induction assumption  $\psi$  is of complexity at most  $k$ . It follows that  $\phi$  is of complexity at most  $k + 1$ .

Let us now demonstrate the converse implication. Assume that  $\phi$  is of complexity at most  $k + 1$ . Let us fix arbitrary  $\beta \in \text{Asn}(\phi)$ . If  $\phi$  is atomic, then it is clearly decided by  $\Gamma(k+1)$ , since atomic formulae are decided already by  $\Gamma(0)$ . Let us assume that  $\phi$  is a compound formula and distinguish cases:

Case 1 Assume that  $\phi$  starts with a negation sign. Let  $\phi = \neg\psi$ . In particular,  $\text{Compl}(\psi) \leq k$ , hence  $\Gamma(k)$  decides  $\psi$ . In such case  $\beta \in \text{Asn}(\psi)$ . We have

$$\psi[\beta] \in \Gamma(k) \vee \neg\psi[\beta] \in \Gamma(k)$$

If the first disjunct holds, then  $\neg\phi[\beta] \in \Gamma(k+1)$  ( $\neg\phi = \neg\neg\psi$ ). If the second one, then  $\phi[\beta] \in \Gamma(k+1)$ , ( $\phi = \neg\psi$ ) since  $\Gamma(k)$  is monotone in  $k$ .

Case 2 Assume that for some set of sentences  $c$ ,  $\phi = \bigvee^* c$ . Then every  $\theta \in c$  is of complexity at most  $k$ , hence  $\Gamma(k)$  decides every disjunct of  $\phi$ . It means that we have

$$\forall \theta \in c \text{ tot}(\theta[\beta], \Gamma(k))$$

Of course  $\beta \in \text{Asn}(\theta)$ , for every  $\theta \in c$ . Now, either for all  $\theta \in c$   $\neg\theta[\beta] \in \Gamma(k)$  or there is  $\theta \in c$  such that  $\theta[\beta] \in \Gamma(k)$ . In the first case  $\neg\bigvee_{\psi \in c} \psi[\beta] \in \Gamma(k+1)$ , in the second  $\bigvee_{\psi \in c} \psi[\beta] \in \Gamma(k)$ .

Case 3 Assume that  $\phi = \exists v\psi$  and  $\psi$  is of complexity at most  $k$ . By our assumption  $\Gamma(k)$  decides  $\psi$  i.e. we have

$$\forall \beta \in \text{Asn}(\psi) (\psi[\beta] \in \Gamma(k) \vee \neg\psi[\beta] \in \Gamma(k))$$

Let us fix arbitrary  $\beta \in \text{Asn}(\phi)$ . By the above and the fact that  $\psi[\beta]$  has at most one free variable we have

$$\forall x (\psi[\beta](\underline{x}) \in \Gamma(k) \vee \neg\psi[\beta](\underline{x}) \in \Gamma(k))$$

i.e.  $\text{tot}_v(\psi[\beta], \Gamma(k))$ . Hence, either

$$\text{tot}_v(\psi[\beta], \Gamma(k)) \wedge \exists x (\psi[\beta](\underline{x}))$$

or

$$\forall x (\neg\psi[\beta](x))$$

holds. By the definition of  $\Gamma(k+1)$  it is now clear that either  $\exists v\psi[\beta] \in \Gamma(k+1)$  or  $\neg\exists v\psi[\beta] \in \Gamma(k+1)$ . Since  $\beta$  was arbitrary this step is finished, and so is the whole proof.  $\square$

**Corollary 215.** *Let  $\mathcal{M} \models \text{PA}$  and  $\phi \in \text{Form}_{\mathcal{M}}$ . For every  $n$  we have*

$$\exists\beta (\phi[\beta] \in \Gamma(n) \vee \neg\phi[\beta] \in \Gamma(n)) \iff \text{Compl}^*(\phi) \leq n$$

*Proof.* Let us fix  $\mathcal{M}$  and  $\phi \in \text{Form}_{\mathcal{M}}$ . The right-to-left implication is an immediate corollary from Lemma 214. Let us assume that for some  $n$  and some  $\beta \in \text{Asn}(\phi)$  we have

$$(\phi[\beta] \in \Gamma(n) \vee \neg\phi[\beta] \in \Gamma(n))$$

Since  $\phi[\beta]$  is a sentence, then for any  $\gamma \in \text{Asn}(\phi[\beta])$  we have

$$\phi[\beta][\gamma] = \phi[\beta].$$

It follows that  $\Gamma(n)$  decides  $\phi[\beta]$ , hence by the left-to-right implication in Lemma 214 we have that the complexity of  $\phi[\beta]$  is at most  $n$ . But the complexity of a formula does not depend on the terms occurring in it, hence the complexity of  $\phi$  is also at most  $n$ .  $\square$

In the following considerations  $\mathcal{M}$  is an arbitrary model of PA. Proof of the next proposition is straightforward:

**Proposition 216.** *Let  $\phi \in \text{Form}_{\mathcal{M}}$ .  $\Gamma_{\text{DC}}^{\mathcal{M}}(\omega)$  decides  $\phi$  if and only if for some  $n$   $\Gamma_{\text{DC}}^{\mathcal{M}}(n)$  decides  $\phi$ .*

*Sketch of the proof.* The implication from right to left follows from monotonicity of  $\Gamma_{\text{DC}}^{\mathcal{M}}(\alpha)$ . From left to right we use Corollary 215: if  $\Gamma_{\text{DC}}^{\mathcal{M}}(\alpha)$  decides  $\phi$ , then for some  $n$  and some  $\beta \in \text{Asn}(\phi)$  either  $\phi[\beta]$  or  $\neg\phi[\beta]$  is an element of  $\Gamma_{\text{DC}}^{\mathcal{M}}(n)$ . Fixing  $n$ , by Corollary 215, the complexity of  $\phi$  is at most  $n$  and, consequently, (once again invoking the corollary)  $\Gamma(n)$  decides  $\phi$ .  $\square$

**Definition 217.** Let  $c \in \text{SetSent}_{\mathcal{M}}$ . We say that  $\Gamma(\alpha)$  *decides*  $c$  if and only if for every  $\psi \in^{\mathcal{M}} c$ ,  $\Gamma(\alpha)$  decides  $\psi$ .

The following proposition is a crucial application of Lemma 214:

**Proposition 218.** *For every  $c \in \text{SetSent}_{\mathcal{M}}$ ,  $\Gamma(\omega)$  decides  $c$  if and only if for some  $n$   $\Gamma(n)$  decides  $c$ .*

*Proof.* The proof of right-to-left implication is straightforward and follows from monotonicity of  $\Gamma$  and the definition of  $\Gamma(\omega)$ . Let us fix  $c \in \text{SetSent}_{\mathcal{M}}$  and suppose that for every  $n \in \omega$ ,  $\Gamma(n)$  does not decide  $c$ . Hence, for every  $n \in \omega$  there exists a sentence  $\psi_n \in^{\mathcal{M}} c$  such that  $\Gamma(n)$  does not decide  $\psi_n$ . By Lemma 214 every  $\psi_n$  is of complexity *strictly greater* than  $n$ . Now, we use induction in  $\mathcal{M}$ : since  $\text{Compl}^*$  is definable in  $\mathcal{M}$ , we have

$$\text{for all } n \in \omega, \mathcal{M} \models \exists\psi \in c (\text{Compl}^*(\psi) > n) \quad (5.12)$$

By the Overspill Lemma (see Lemma 54) there exists  $d \in M$  such that for all  $n \in \omega$ ,  $d >^{\mathcal{M}} n$  and

$$\mathcal{M} \models \exists\psi \in c (\text{Compl}^*(\psi) > d) \quad (5.13)$$

In particular, by Lemma 214, there exists  $\psi \in^{\mathcal{M}} c$  such that for no  $n$ ,  $\Gamma(n)$  decides  $\psi$ . Hence,  $\Gamma(\omega)$  does not decide  $\psi$  and, consequently,  $\Gamma(\omega)$  does not decide  $c$ .  $\square$

Now we can prove Proposition 211:

*Proof of Proposition 211.* The inclusion  $\Gamma(\omega) \subseteq \Gamma(\Gamma(\omega))$  follows from the monotonicity of  $\Gamma$ . To demonstrate the converse inclusion let us fix  $\phi$  and assume that  $\phi \in \Gamma(\Gamma(\omega))$ . If  $\phi$  is an atomic sentence or negation of an atomic sentence (in the sense of  $\mathcal{M}$ ), then it is decided already by  $\Gamma(0)$  and hence it belongs to  $\Gamma(\Gamma(\omega))$  if and only if it belongs to  $\Gamma(\omega)$ . We shall split the proof in four steps depending on the main connective of  $\phi$ . The steps will jointly complete our reasoning:

Case 1. Let us suppose that  $\phi =^{\mathcal{M}} \bigvee^* c$  for some  $c \in M$ . Since  $\phi \in \Gamma(\Gamma(\omega))$ , then for every  $\psi \in^{\mathcal{M}} c$  we have

$$\text{tot}(\psi, \Gamma(\omega))$$

and for some  $\psi \in c, n \in \omega, \psi \in \Gamma(n)$ . By Proposition 218 there is  $k \in \omega$  such that  $\Gamma(k)$  decides  $c$ . In particular, it must be the case that  $\phi \in \Gamma(k+1)$ , since  $\Gamma(k)$  decides  $\psi$  and  $\psi \in \Gamma(n)$  (otherwise  $\Gamma(\omega)$  would be inconsistent). Hence,  $\phi \in \Gamma(\omega)$  as well.

Case 2. Let us suppose that  $\phi =^{\mathcal{M}} \neg \bigvee^* c$  for some  $c \in M$ . Since  $\phi \in \Gamma(\Gamma(\omega))$ , then for every  $\psi \in^{\mathcal{M}} c$  we have

$$\neg \psi \in \Gamma(\omega)$$

In particular,  $c$  is decided by  $\Gamma(\omega)$ . By Proposition 218 there is  $k \in \omega$  such that  $\Gamma(k)$  decides  $c$ . Since  $\Gamma(\omega)$  is consistent, it means that for every  $\psi \in^{\mathcal{M}} c, \neg \psi \in \Gamma(k)$ . Hence,  $\phi \in \Gamma(k+1)$  and, consequently,  $\phi \in \Gamma(\omega)$ .

Case 3. Assume  $\phi = \exists v \psi$ . Then it follows that for some  $a \in M, \psi(\underline{a}) \in \Gamma(\omega)$ . Hence, for some  $n, \psi(\underline{a}) \in \Gamma(n)$ . Let us fix  $n$ . By Corollary 215 and Lemma 214,  $\Gamma(n)$  decides  $\psi$ . In particular, for the chosen  $n \exists v \psi \in \Gamma(n)$  and hence  $\phi \in \Gamma(\omega)$ .

Case 4. Assume  $\phi = \neg \exists v \psi$ . Then it follows that for every  $a \in M, \neg \psi(\underline{a}) \in \Gamma(\omega)$ . In particular, for some  $n, \neg \psi(\underline{0}) \in \Gamma(n)$ . Hence, by Lemma 214  $\psi$  is of complexity at most  $n$  and as such is decided by  $\Gamma(n)$ . Let us fix  $n$ . It follows that for every  $a \in M, \neg \psi(\underline{a}) \in \Gamma(n)$ , for otherwise  $\Gamma(\omega)$  would be inconsistent. In particular,  $\phi \in \Gamma(n+1)$  and hence also  $\phi \in \Gamma(\omega)$ .  $\square$

By the above proposition we know that for arbitrary  $\mathcal{M} \models \text{PA}$ ,

$$(\mathcal{M}, \Gamma_{\text{DC}}^{\mathcal{M}}(\omega)) \models \text{WPT}^- + \text{DC}.$$

Let us now show that such models satisfy also internal induction, which will complete the proof of model-theoretical conservativity of  $\text{WPT}^- + \text{DC} + \text{INT}$ .

**Proposition 219.** *For every  $\mathcal{M} \models \text{PA}$ ,  $(\mathcal{M}, \Gamma_{\text{DC}}^{\mathcal{M}}(\omega)) \models \text{INT}$ .*

*Proof.* Let us fix  $\mathcal{M}$  and  $\phi \in \text{Form}_{\mathcal{M}}^{\leq 1}$ . Assume that  $\phi(\underline{0}) \in \Gamma(\omega)$  and for every  $a \in M$ , if  $\phi(\underline{a}) \in \Gamma(\omega)$ , then  $\phi(\underline{a+1}) \in \Gamma(\omega)$ . By our assumption there is an  $n \in \omega$  such that  $\phi(\underline{0}) \in \Gamma(n)$ . Let us fix  $n$ . Since the complexity of a formula does not depend on which terms occur in it, by Lemma 214 we can conclude that  $\phi$  is of complexity at most  $n$ . In consequence,  $\Gamma(n)$  decides  $\phi$  and we have

$$\text{for every } a \in M \quad \phi(\underline{a}) \in \Gamma(n) \iff \phi(\underline{a}) \in \Gamma(\omega)$$

Hence, for every  $a \in M$ , if  $\phi(\underline{a}) \in \Gamma(n)$ , then  $\phi(\underline{a+1}) \in \Gamma(n)$ . Since  $\Gamma(n)$  is definable in  $\mathcal{M}$ , then

$$\theta(x) := \phi(\underline{x}) \in \Gamma(n)$$

is an arithmetical formula and we have

$$\mathcal{M} \models \theta(0) \wedge \forall x(\theta(x) \rightarrow \theta(x+1))$$

Hence, by induction in  $\mathcal{M}$  we get  $\mathcal{M} \models \forall x\theta(x)$  and in particular, for every  $a \in M$ ,  $\phi(\underline{a}) \in \Gamma(n)$ . Hence, also for every  $a \in M$ ,  $\phi(\underline{a}) \in \Gamma(\omega)$  and our proof is finished.  $\square$

### Digression 2: Comparing Weak and Strong Kleene Logic

One of the most interesting questions in axiomatic truth theories is whether compositional theories based on Weak Kleene Logic are *weaker* than the respective compositional theories based on Strong Kleene Logic. The question was originally posed by Volker Halbach (personal communication) in the context of  $\text{WKF}^-$  and  $\text{KF}^-$  and used the notion of relative truth definability as the explication of the notion of "strength". More concretely, the question was:

Is  $\text{KF}^-$  relatively truth definable in  $\text{WKF}^-$ ?

Fujimoto (in [17];  $\text{WKF}^-$  was introduced there) showed that the converse holds. In particular, in the context of typed theories of truth we may ask the following variant:

Is  $\text{PT}^-$  relatively truth definable in  $\text{WPT}^-$ ?

Also, in this case one can demonstrate that  $\text{WPT}^-$  is Fujimoto definable in  $\text{PT}^-$ . Answers to both questions about the opposite direction are yet to be found. What we were able to discover so far is that, in some contexts, applying the Weak Kleene Logic instead of the Strong one results in weaker theory: we have shown that  $\text{WPT}^- + \text{DC} + \text{INT}$  is much weaker than  $\text{PT}^- + \text{DC} + \text{INT}$ ; in particular,

**Theorem 220.**  $\text{PT}^- + \text{DC} + \text{INT}$  is not relatively truth definable in  $\text{WPT}^- + \text{DC} + \text{INT}$ .

This proposition follows easily from the proof-theoretical non-conservativity of the former theory and model-theoretical conservativity of the latter (jointly with Proposition 85). In fact, the above theorem may be strengthened: even  $\text{PT}^- + \text{INT}(\text{tot})$  is stronger than  $\text{WPT}^- + \text{DC} + \text{INT}$  (although not as much as  $\text{PT}^- + \text{DC} + \text{INT}$ ): as was shown by Bartosz Wcisło (Theorem 2.1 in [5]) adding to  $\text{PT}^-$  internal induction only for total formulae suffices to yield a model-theoretically non-conservative theory (over PA).<sup>3</sup> Consequently,  $\text{PT}^- + \text{INT}(\text{tot})$  is not relatively truth definable in  $\text{WPT}^- + \text{DC} + \text{INT}$ .

<sup>3</sup> It can be shown that every model of  $\text{PT}^- + \text{INT}(\text{tot})$  is short recursively saturated.

### 5.3 Reflection principles

In this subsection we shall study the strength of various reflection principles when added to  $PT^-$  and  $WPT^-$ . Since our primary objective is to verify whether "Many Faces" theorem can be transferred to the partially compositional setting we shall work solely with truth theories (as not all equivalences in this theorem are known for  $CS^-$ ). Let us make one preliminary observation that, from the very beginning simplifies our considerations. Let us recall that

$$\text{Pr}_{\text{CPC}}(x)$$

is an arithmetical predicate representing in PA the property

" $x$  is an instantiation of formula provable in CPC with formulae of  $\mathcal{L}_{\text{PA}}$ ."

(where CPC stands for Classical *Propositional* Calculus).

**Proposition 221.**  $PT^- + \forall\phi(\text{Pr}_{\text{CPC}}(\phi) \rightarrow T(\phi)) \vdash \text{Tot}$ . *The same holds for  $WPT^-$ .*

*Proof.* The proof is immediate: since (PA proves that) for each  $\phi \in \text{Sent}_{\mathcal{L}_{\text{PA}}}$ ,  $\phi \vee \neg\phi$  is an instantiation of a CPC provable formula, then both in  $PT^-$  as well as in  $WPT^-$  case we get

$$PT^- + \forall\phi(\text{Pr}_{\text{CPC}}(\phi) \rightarrow T(\phi)) \vdash \forall\phi T(\phi \vee \neg\phi)$$

the same being true for  $WPT^-$ . Since both theories prove

$$\forall\phi(T(\phi \vee \neg\phi) \rightarrow (T(\phi) \vee T(\neg\phi)))$$

our proof is finished. □

In particular, by Proposition 105 we see that if  $\Theta$  is an arbitrary reflection principle (of type considered in Section 3.4, Chapter 3), then

$$PT^- + \Theta \text{ and } WPT^- + \Theta$$

are deductively equivalent. Hence, without loss of generality in our considerations we can focus exclusively on  $PT^-$ . Let us make the following easy observation that will however bring to light very interesting consequences.

**Proposition 222.**  $PT^- + \forall\phi(\text{Pr}_{\text{CPC}}^T(\phi) \rightarrow T(\phi)) \vdash \text{CT}_0$

*Proof.* By the "Many Faces" theorem together with Propositions 221 and 105 all we need to show is that

$$PT^- + \forall\phi(\text{Pr}_{\text{CPC}}^T(\phi) \rightarrow T(\phi)) \vdash \text{Cons}$$

The above is in fact very easy: working in  $PT^- + \forall\phi(\text{Pr}_{\text{CPC}}^T(\phi) \rightarrow T(\phi))$  assume there exists  $\phi$  such that

$$T(\phi) \wedge T(\neg\phi)$$

then we have  $\text{Pr}_{\text{CPC}}^T(\phi \wedge \neg\phi)$ . Since also, by *ex falso quodlibet*, we have

$$\text{Pr}_{\text{CPC}}^T((\phi \wedge \neg\phi) \rightarrow (0 = 1)),$$

we can invoke *modus ponens* inside  $\text{Pr}_{\text{CPC}}^T(x)$  and obtain  $\text{Pr}_{\text{CPC}}^T(0 = 1)$ . Applying the closure reflection principle we obtain

$$T(0 = 1)$$

and the just obtained contradiction ends our proof.  $\square$

In particular, we can reconstruct this part of "Many Faces" theorem, which refers to closure reflection principles:

**Theorem 223.** *The following theories have the same consequences:*

1.  $\text{PT}^- + \forall\phi(\text{Pr}_{\text{CPC}}^T(\phi) \rightarrow T(\phi))$
2.  $\text{PT}^- + \forall\phi(\text{Pr}_{\emptyset}^T(\phi) \rightarrow T(\phi))$
3.  $\text{PT}^- + \forall\phi(\text{Pr}_{\text{PA}}^T(\phi) \rightarrow T(\phi))$
4.  $\text{CT}_0$

Let us observe that, by the results of the last section, any above-listed theory is deductively equivalent to  $\text{PT}_0$  and is a notational variant of  $\text{WPT}_0^{++}$ . In particular, its set of arithmetical consequences is deductively equivalent to the first reflective closure of  $\text{PA}$ ,  $\mathcal{UR}^\omega(\text{PA})$ , as defined in Section 3.4. Let us now turn to the *completeness* reflection principles, as here the differences emerge. To begin with, we shall show the obvious lower bound on the arithmetical strength of the resulting extensions of  $\text{PT}^-$ . Since our proofs are, in a sense, uniform in the *theory* we take the uniform reflection over, we will formulate the following theorems in greater generality. In what follows,  $\text{Th}$  is any arithmetically definable arithmetical theory (see Definition 7).

**Proposition 224.**  $\text{PT}^- + \forall\phi (\text{Pr}_{\text{Th}}(\phi) \rightarrow T(\phi)) \vdash \mathcal{UR}(\text{Th})$

*Proof.* Let us fix arbitrary formula  $\phi(x_0, \dots, x_n)$ . We work in  $\text{PT}^- + \forall\phi (\text{Pr}_{\text{Th}}(\phi) \rightarrow T(\phi))$ : for arbitrary  $x_0, \dots, x_n$ , by the reflection principle we have

$$\text{Pr}_{\text{Th}}(\phi(\underline{x}_0, \dots, \underline{x}_n)) \rightarrow T(\phi(x_0, \dots, x_n))$$

Hence, since  $\text{PT}^- \vdash \text{UTB}^-$ , the succedent of the above implication is equivalent to  $\phi(x_0, \dots, x_n)$  and we get

$$\text{Pr}_{\text{Th}}(\phi(\underline{x}_0, \dots, \underline{x}_n)) \rightarrow \phi(x_0, \dots, x_n)$$

which ends the proof.  $\square$

It turns out that, contrary to the  $\text{CT}^-$  case<sup>4</sup>, in the contexts where  $\text{PA} + \mathcal{UR}(\text{Th}) \vdash \text{SCon}(\text{Th})$  (see Proposition 72),  $\text{PT}^-$  does not prove any more than the above proposition shows.

<sup>4</sup> By "Many Faces" Theorem  $\text{CT}^- + \forall\phi (\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi))$  is deductively equivalent to  $\text{CT}_0$ , hence arithmetically much stronger than  $\mathcal{UR}(\text{PA})$ .

**Theorem 225.**  $\text{PT}^- + \forall\phi (\text{Pr}_{\text{Th}}(\phi) \rightarrow T(\phi))$  is proof-theoretically conservative over  $\text{PA} + \text{SCon}(\text{Th})$ .

*Proof.* We shall demonstrate that for each model  $\mathcal{K} \models \text{PA} + \text{SCon}(\text{Th})$  there exists a model  $\mathcal{K}' \models \text{PT}^- + \forall\phi (\text{Pr}_{\text{Th}}(\phi) \rightarrow T(\phi))$  which is  $\mathcal{L}_{\text{PA}}$ -elementarily equivalent to  $\mathcal{K}$ . If  $\text{PA} + \text{SCon}(\text{Th})$  is inconsistent, then there is nothing to prove, so let us fix  $\mathcal{K} \models \text{PA} + \text{SCon}(\text{Th})$ . By Proposition 67 there exists  $\mathcal{M} \models \text{UTB} + \text{SCon}(\text{Th})$  such that  $\mathcal{K} \preceq_{\mathcal{L}_{\text{PA}}} \mathcal{M}$ . Then, by Theorem 66,  $\mathcal{M}$  is recursively saturated. We shall use the fact that in such models there are particularly simple extensions of  $\text{PT}^-$  truth predicate. Intuitively their simplicity rests upon the fact that each their approximation is *definable* (in fact we used this property while proving Theorem 204 and Theorem 209). Let  $\Gamma^{\mathcal{M}}$  be the operator defined in 115. By Lemma 117  $\Gamma^{\mathcal{M}}(\omega) \subseteq M$  is a fixpoint of  $\Gamma^{\mathcal{M}}$  and hence an extension for the  $\text{PT}^-$  truth predicate. Let us define (as we keep  $\mathcal{M}$  fixed we shall skip the superscript in  $\Gamma$ ):

$$\Gamma(\mathcal{m}) := \{\phi \in M \mid \neg\phi \notin \Gamma(\omega)\} \quad (5.14)$$

Now, by Proposition 116,  $(\mathcal{M}, \Gamma(\mathcal{m})) \models \text{PT}^-$ . We shall show that

$$(\mathcal{M}, \Gamma(\mathcal{m})) \models \forall\phi (\text{Pr}_{\text{Th}}(\phi) \rightarrow T(\phi))$$

which will end the whole proof. Let us fix arbitrary sentence  $\phi$  in the sense of  $\mathcal{M}$  and assume that

$$\neg T(\phi)$$

This implies that  $\ulcorner \neg\phi \urcorner \in \Gamma(\omega)$ . Hence, there is an  $n$  such that  $\neg\phi \in \Gamma(n)$ . Now, as we already observed in Lemma 205, for every  $n$  there exists an arithmetical formula  $x \in \Gamma(n)$  such that

$$x \in^{\mathcal{M}} \Gamma(n) = \Gamma(n)$$

In particular, we have

$$\mathcal{M} \models \neg\phi \in \Gamma(n)$$

Let  $\mathcal{N}$  be any model satisfying the thesis of Theorem 70. Then, since  $\mathcal{M} \prec \mathcal{N}$  and  $\neg\phi \in M$  then

$$\mathcal{N} \models \neg\phi \in \Gamma(n)$$

and the same is true also in  $\mathcal{M}$ , i.e.

$$\mathcal{M} \models (\mathcal{N} \models_{\mathcal{N}} \neg\phi \in \Gamma(n)) \quad (5.15)$$

Now by the external induction on  $k \in \omega$  in the (ZFC) formula

$$\Theta(k) := \left( \mathcal{M} \models \forall\psi(\bar{v}) \left( \mathcal{N} \models_{\mathcal{N}} \forall\bar{v} ((\psi(\bar{v}) \in \Gamma(k)) \rightarrow \psi(\bar{v})) \right) \right) \quad (5.16)$$

we show that

$$\forall k \Theta(k)$$

which, using 5.15 and the fact that  $\neg\phi$  is a sentence, will end the proof, since we will get

$$\mathcal{M} \models (\mathcal{N} \models_{\mathcal{N}} \neg\phi)$$

which implies  $\mathcal{M} \models \neg \text{Pr}_{\text{Th}}(\phi)$  by the Arithmetised Completeness Theorem (Theorem 49). Let us decipher the meaning of  $\Theta(k)$ : in  $\mathcal{M}$  it has to be satisfied that for every formula  $\psi$  (i.e. every formula *in the sense of*  $\mathcal{M}$ ) in  $\mathcal{N}$  the one-way uniform Tarski biconditional hold for  $\psi$ . In particular,  $\forall \bar{v}$  is meant to bind all the free variables in  $\psi$  and  $\psi(\bar{v})$  denotes the unique formula (in the sense of  $\mathcal{M}$ ) of the following kind

$$\exists w_{i_0} \dots \exists w_{i_k} \left( \left( \bigwedge_{j \leq k} w_{i_j} = \underline{v_{i_j}} \right) \wedge \psi[w_{i_0}/v_{i_0}, \dots, w_{i_k}/v_{i_k}] \right).$$

Let us note that this time the (possibly non-standard) quantifier prefixes need not trouble us: since

$$\mathcal{N} \models_{\mathcal{N}} \forall \bar{v} \left( (\psi(\bar{v}) \in \Gamma(k)) \rightarrow \psi(\bar{v}) \right)$$

is an  $\mathcal{L}_{\text{PA}}$  formula (possibly with parameters) for a fixed  $k$  the argument can be carried out in PA. The base step for  $k = 0$  follows easily, because in such a case ( $\mathcal{M}$  satisfies that) any  $\psi$  such that ( $\mathcal{M}$  satisfies that  $\mathcal{N}$  satisfies that) for some  $\bar{x}$

$$\psi(\bar{x}) \in \Gamma(0)$$

has to be an atomic formula of type  $s = t$ , for some (not necessarily closed) terms  $s, t$  ( $s, t$  are terms *in the sense of*  $\mathcal{M}$ .) By the definition of  $\Gamma(0)$

$$\mathcal{M} \models (\mathcal{N} \models_{\mathcal{N}} (s(\bar{x})^\circ = t(\bar{x})^\circ))$$

By the PA provable properties of satisfaction relation and the fact that for every  $x$ ,  $\underline{x}^\circ = x$ , the above is equivalent to

$$\mathcal{M} \models (\mathcal{N} \models_{\mathcal{N}} (s(\bar{x}) = t(\bar{x}))).$$

In the proof of the induction step we use the fact that each  $\Gamma(k)$  is defined by induction on the build-up of  $\phi$  in terms of  $\Gamma(k-1)$ . We distinguish cases (and all the time reason in  $\mathcal{M}$ ).

*Case 1:* Assume  $\psi(\bar{x}) = \neg(\psi_1(\bar{x}) \vee \psi_2(\bar{x}))$  then we have for arbitrary  $\bar{a} \in N$ :

$$\begin{aligned} \mathcal{N} \models_{\mathcal{N}} (\neg\psi(\bar{a}) \in \Gamma(k)) &\iff \mathcal{N} \models_{\mathcal{N}} (\neg\psi_1(\bar{a}) \in \Gamma(k-1) \wedge \neg\psi_2(\bar{a}) \in \Gamma(k-1)) \\ &\iff \mathcal{N} \models_{\mathcal{N}} \neg\psi_1(\bar{a}) \in \Gamma(k-1) \wedge \mathcal{N} \models_{\mathcal{N}} \neg\psi_2(\bar{a}) \in \Gamma(k-1) \\ &\iff \mathcal{N} \models_{\mathcal{N}} \neg\psi_1(\bar{a}) \wedge \mathcal{N} \models_{\mathcal{N}} \neg\psi_2(\bar{a}) \\ &\implies \mathcal{N} \models_{\mathcal{N}} (\neg\psi_1 \wedge \neg\psi_2)(\bar{a}) \\ &\iff \mathcal{N} \models_{\mathcal{N}} \neg(\psi_1 \vee \psi_2)(\bar{a}) \end{aligned}$$

*Case 2:* Assume  $\psi(\bar{x}) = (\psi_1(\bar{x}) \vee \psi_2(\bar{x}))$  then we have for arbitrary  $\bar{a} \in N$ :

$$\begin{aligned} \mathcal{N} \models_{\mathcal{N}} (\psi(\bar{a}) \in \Gamma(k)) &\iff \mathcal{N} \models_{\mathcal{N}} (\psi_1(\bar{a}) \in \Gamma(k-1) \vee \psi_2(\bar{a}) \in \Gamma(k-1)) \\ &\iff \mathcal{N} \models_{\mathcal{N}} \psi_1(\bar{a}) \in \Gamma(k-1) \vee \mathcal{N} \models_{\mathcal{N}} \psi_2(\bar{a}) \in \Gamma(k-1) \\ &\implies \mathcal{N} \models_{\mathcal{N}} \psi_1(\bar{a}) \vee \mathcal{N} \models_{\mathcal{N}} \psi_2(\bar{a}) \\ &\iff \mathcal{N} \models_{\mathcal{N}} (\psi_1 \vee \psi_2)(\bar{a}) \end{aligned}$$

Case 3: Assume  $\psi(\bar{x}) = \exists y\psi_1(\bar{x})$ , then for arbitrary  $\bar{a} \in N$  we have

$$\begin{aligned} \mathcal{N} \models_{\mathcal{N}} (\psi(\bar{a}) \in \Gamma(k)) &\iff \mathcal{N} \models_{\mathcal{N}} (\exists y(\psi_1(\bar{a}, y) \in \Gamma(k-1))) \\ &\iff \exists y \in N\mathcal{N} \models_{\mathcal{N}} \psi_1(\bar{a}, y) \in \Gamma(k-1) \\ &\implies \exists y \in N\mathcal{N} \models_{\mathcal{N}} \psi_1(\bar{a})(y) \\ &\iff \mathcal{N} \models_{\mathcal{N}} \exists y\psi_1(\bar{a}) \end{aligned}$$

Case 4: Assume  $\psi(\bar{x}) = \neg\exists y\psi_1(\bar{x})$ , then for arbitrary  $\bar{a} \in N$  we have

$$\begin{aligned} \mathcal{N} \models_{\mathcal{N}} (\psi(\bar{a}) \in \Gamma(k)) &\iff \mathcal{N} \models_{\mathcal{N}} (\forall y(\neg\psi_1(\bar{a}, y) \in \Gamma(k-1))) \\ &\iff \forall y \in N\mathcal{N} \models_{\mathcal{N}} \neg\psi_1(\bar{a}, y) \in \Gamma(k-1) \\ &\implies \forall y \in N\mathcal{N} \models_{\mathcal{N}} \neg\psi_1(\bar{a}, y) \\ &\iff \forall y \in N\mathcal{N} \models \forall y\neg\psi_1(\bar{a}) \\ &\iff \mathcal{N} \models \neg\exists y\psi_1(\bar{a}) \end{aligned}$$

This ends the whole proof. □

Let us note two particularly interesting corollaries

**Corollary 226.**  $\text{PT}^- + \forall\phi (\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi))$  is conservative over  $\mathcal{UR}(\text{PA})$ .

*Proof.* This easily follows from the above theorem and Proposition 72. □

**Corollary 227.**  $\text{PT}^- + \forall\phi (\text{Pr}_{\emptyset}(\phi) \rightarrow T(\phi))$  is conservative over  $\text{PA}$ .

*Proof.* This is an easy consequence of the above theorem, Proposition 72 and Theorem 62. □

## 6. SUMMARY: THE BIG PICTURE

The main aim of our research in this dissertation was to determine the proof-theoretical strength of extensions of three basic compositional theories of truth:  $CT^-$ ,  $PT^-$  and  $WPT^-$ . We studied which arithmetical sentences becomes provable when we add principles of the following three kinds to these theories:

1.  $\Delta_0$  induction for  $\mathcal{L}_T$ ;
2. Reflection Principles, which can be subdivided into
  - (a) Completeness Reflection Principles:
    - i. Global Reflection:  $\forall\phi \ (\text{Pr}_{\text{PA}}(\phi) \rightarrow T(\phi))$ ;
    - ii. First-Order Logic Completeness Principle:  $\forall\phi \ (\text{Pr}_\emptyset(\phi) \rightarrow T(\phi))$ ;
    - iii. Propositional Logic Completeness Principle:  $\forall\phi \ (\text{Pr}_{\text{CPC}}(\phi) \rightarrow T(\phi))$ ;
  - (b) Closure Reflection Principles:
    - i. First-Order Logic Closure Principle:  $\forall\phi \ (\text{Pr}_\emptyset^T(\phi) \rightarrow T(\phi))$ ;
    - ii. Propositional Logic Closure Principle:  $\forall\phi \ (\text{Pr}_{\text{CPC}}^T(\phi) \rightarrow T(\phi))$ ;
3. Additional axioms:
  - (a) Disjunctive Correctness (for every basic compositional truth theory defined in a slightly different way, see Definition 134);
  - (b) Internal Induction:  $\forall\phi \ \left( \left( T(\phi(0)) \wedge \forall x \ (T(\phi(x)) \rightarrow T(\phi(x+1))) \right) \rightarrow \forall x T(\phi(x)) \right)$ .

The first of the original results of this dissertation was presented in Chapter 4. It stated that when  $\Delta_0$  induction for  $\mathcal{L}_T$  is added to  $CT^-$ , then the resulting theory proves the Global Reflection (we proved a slightly stronger statement, that Global Reflection is derivable in the respective theory of satisfaction ( $CS_0$ )). Our contribution consists of directly fixing the gap in Kotlarski's proof from [28]. The gap was discovered in 2008. Jointly with the results of Cieśliński and Enayat, this allows us to conclude that extending  $CT^-$  with *any* principle of the above list results in the same theory (up to deductive equivalence). This is the content of "Many Faces" Theorem 170. In Theorem 126, with the major help of Kotlarski's earlier result, we characterized the set of arithmetical consequences of  $CT_0$  as theorems of  $\mathcal{UR}^\omega(\text{PA})$ .

Having realised that, over  $CT^-$ , many intuitively different principles yield the same theory, in Chapter 5, we studied whether this phenomenon transfers to the setting of non-classically compositional truth theories. Below, we summarize the findings of this chapter:

1.  $\Delta_0$  Induction:

It was shown that extending  $PT^-$  with  $\Delta_0$  induction for  $\mathcal{L}_T$  results in the theory deductively equivalent to  $CT_0$ . The problem for the analogous extension of  $WPT^-$  is open. However, we showed that if  $WPT^-$  is extended with compositional axioms for bounded quantifiers (which are added to the arithmetised language as new primitive symbols; we called this theory  $WPT^-_+$ ) or strong implication (which is added to the arithmetised language as a new primitive symbol; we called this theory  $FPT^-$ ), then the  $\Delta_0$  inductive extensions of these theories are "the same" as  $CT_0$ , modulo the translation of arithmetised languages. To formalise this relation, we introduced the notion of a *notational variant of a theory of truth*; hence, proving that  $FPT_0$ ,  $WPT_0^+$  and  $CT_0$  are mutual notational variants of each other. We showed that all these results transfer to the setting of theories of satisfaction.

## 2. Reflection Principles

We started by observing that if  $\phi$  is any reflection principle listed in point 2 of the above list, then  $PT^- + \phi$  is deductively equivalent to  $WPT^- + \phi$ ; hence, without loss of generality, we may concentrate on the extensions of  $PT^-$ . We showed that, over  $PT^-$ , closure reflection principles are strictly stronger, than their "completeness" analogues. More concretely: if  $\phi$  is any closure reflection principle, then  $PT^- + \phi$  is deductively equivalent to  $CT_0$ . However,

- (a) the arithmetical consequences of  $PT^- + \forall\phi (Pr_{PA}(\phi) \rightarrow T(\phi))$  are the same as consequences of  $\mathcal{UR}(PA)$ . Hence,  $PT^- + \forall\phi (Pr_{PA}(\phi) \rightarrow T(\phi))$  is proof-theoretically weaker than  $CT_0$ ;
- (b)  $PT^- + \forall\phi (Pr_{\emptyset}(\phi) \rightarrow T(\phi))$  is proof-theoretically conservative over PA.

## 3. Additional Axioms:

We have shown that  $PT^- + DC + INT$  is a one more theory of truth deductively equivalent to  $CT_0$ . However,

- (a)  $WPT^- + DC + INT$  is model-theoretically conservative over PA (we remind that in this case the disjunctive correctness axioms are adjusted to Weak Kleene logic).
- (b)  $PT^- + DC + INT(\text{tot})$  is proof-theoretically conservative over PA.

In particular, both the above theories are strictly weaker than  $PT^- + DC + INT$ .

Let us end this dissertation by pointing to possible lines of continuation of our work. We ask the following questions:

Question 1 What is the proof-theoretical strength of  $CT^- + DC$ ?

Question 2 What is the proof-theoretical strength of  $CT^- + \forall\phi (Pr_{CPC}(\phi) \rightarrow T(\phi))$ ; i.e.  $CT^-$  extended with the completeness principle "All tautologies of Propositional Logic are true"?

Question 3 Is  $WPT_0$  deductively equivalent to  $CT_0$  (as  $PT_0$  and  $WPT_0^+$  are)? If not, is it proof-theoretically conservative over PA?

## BIBLIOGRAPHY

- [1] Lev D. Beklemishev. Reflection principles and provability algebras in formal arithmetic. *Russian Mathematical Surveys*, 60(2):197, 2005.
- [2] Andrea Cantini. Notes on formal theories of truth. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 35(1):97–130, 1989.
- [3] Andrea Cantini. A theory of formal truth as strong as  $ID_1$ . *The Journal of Symbolic Logic*, 55(1):244–259, 1990.
- [4] Cezary Cieśliński. *The Epistemic Lightness of Truth. Deflationism and its Logic*. Cambridge University Press, forthcoming.
- [5] Cezary Cieśliński, Mateusz Łełyk, and Bartosz Wcisło. Models of  $PT^-$  with internal induction for total formulae. *The Review of Symbolic Logic*, 10(1):187–202, 2017.
- [6] Cezary Cieśliński. Deflationary truth and pathologies. *Journal of Philosophical Logic*, 39(3):325–337, 2010.
- [7] Cezary Cieśliński. Truth, conservativeness, and provability. *Mind*, 119(474):409–422, 2010.
- [8] Walter Dean. Arithmetical reflection and the provability of soundness. *Philosophia Mathematica*, 23(1):31–64, 2014.
- [9] Ali Enayat and Albert Visser. New constructions of satisfaction classes. In Theodora Achourioti, Henri Galinon, José Martínez Fernández, and Kentaro Fujimoto, editors, *Unifying the Philosophy of Truth*. Springer-Verlag, 2015.
- [10] H.B. Enderton. *A Mathematical Introduction to Logic*. Harcourt/Academic Press, 2001.
- [11] Solomon Feferman. Transfinite recursive progressions of axiomatic theories. *Journal of Symbolic Logic*, 27(3):259–316, 09 1962.
- [12] Martin Fischer. Minimal truth and interpretability. *Review of Symbolic Logic*, 2(4):799–815, 2009.
- [13] Martin Fischer. Truth and speed-up. *Review of Symbolic Logic*, 7(2):319–340, 2014.
- [14] Martin Fischer. Deflationism and instrumentalism. In Kentaro Fujimoto, José Martínez Fernández, Henri Galinon, and Theodora Achourioti, editors, *Unifying the Philosophy of Truth*. Springer Netherlands, 2015.
- [15] Martin Fischer and Leon Horsten. The expressive power of truth. *Review of Symbolic Logic*, 8(2):345–369, 2015.

- 
- [16] Torkel Franzen. *Inexhaustibility: an Inexhaustive Treatment*. A K Peters/CRC Press, 2004.
- [17] Kentaro Fujimoto. Relative truth definability of axiomatic truth theories. *Bulletin of Symbolic Logic*, 16(3):305–344, 2010.
- [18] Kentaro Fujimoto. Classes and truths in set theory. *Annals of Pure and Applied Logic*, 163(11):1484–1523, 11 2012.
- [19] Petr Hájek and Pavel Pudlák. *Metamathematics of First-Order Arithmetic*. Springer-Verlag, 1993.
- [20] Volker Halbach. *Axiomatic Theories of Truth*. Cambridge University Press, 2011.
- [21] Richard G. Heck. Consistency and the theory of truth. *Review of Symbolic Logic*, 8(3):424–466, 2015.
- [22] Leon Horsten and Graham E. Leigh. Truth is simple. *Mind*, 126(501):195–232, 2017.
- [23] Matt Kaufmann and James Schmerl. Remarks on weak notions of saturation in models of Peano arithmetic. *Journal of Symbolic Logic*, 52(1):129–148, 1987.
- [24] Richard Kaye. *Models of Peano Arithmetic*. Clarendon Press, 1991.
- [25] Richard Kaye and Tin Lok Wong. On interpretations of arithmetic and set theory. *Notre Dame J. Formal Logic*, 48(4):497–510, 10 2007.
- [26] Jeffrey Ketland. Deflationism and Tarski’s paradise. *Mind*, 108(429):69–94, 1999.
- [27] Roman Kossak and James Schmerl. *The Structure of Models of Peano Arithmetic*. Clarendon Press, 2006.
- [28] Henryk Kotlarski. Bounded induction and satisfaction classes. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 32(31-34):531–544, 1986.
- [29] Henryk Kotlarski, Stanisław Krajewski, and Alistair Lachlan. Construction of satisfaction classes for nonstandard models. *Canadian Mathematical Bulletin*, 24:283–93, 1981.
- [30] Saul A. Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72(19):690–716, 1975.
- [31] Graham E. Leigh. Conservativity for theories of compositional truth via cut elimination. *The Journal of Symbolic Logic*, 80(3):845–865, 2015.
- [32] Graham E. Leigh and Carlo Nicolai. Axiomatic truth, syntax and metatheoretic reasoning. *Review of Symbolic Logic*, 6(4):613–636, 2013.
- [33] Mateusz Łełyk and Bartosz Wcisło. Models of weak theories of truth. *Archive for Mathematical Logic*, 2017. doi:10.1007/s00153-017-0531-1.
- [34] Per Lindström. *Aspects of Incompleteness*. Lecture Notes in Logic. Cambridge University Press, 2017.

- 
- [35] D. Marker. *Model Theory : An Introduction*. Graduate Texts in Mathematics. Springer New York, 2002.
- [36] Roman Murawski. *Funkcje rekurencyjne i elementy metamatematyki. Problemy zupełności, rozstrzygalności, twierdzenia Gödla*. Wydawnictwo UAM, 2010.
- [37] Stephen Neale. *Descriptions*. MIT Press, 1990.
- [38] Pavel Pudlak. The lengths of proofs. In S.R. Buss, editor, *Handbook of Proof Theory*, pages 547–637. Elsevier, 1998.
- [39] Stewart Shapiro. Proof and truth: Through thick and thin. *Journal of Philosophy*, 95(10):493–521, 1998.
- [40] Stephen G. Simpson. Partial realizations of Hilbert’s program. *Journal of Symbolic Logic*, 53(2):349–363, 1988.
- [41] Stephen G. Simpson. *Subsystems of Second-Order Arithmetic*. Springer Verlag, 1999.
- [42] Andrea Stollo. Deflationism and the invisible power of truth. *Dialectica*, 67(4):521–543, 2013.
- [43] W. W. Tait. Finitism. *Journal of Philosophy*, 78(9):524–546, 1981.
- [44] Alfred Tarski. The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research*, 4(3):341–376, 1943.
- [45] Neil Tennant. Deflationism and the Gödel phenomena. *Mind*, 111(443):551–582, 2002.
- [46] Bartosz Wcisło and Mateusz Łełyk. Notes on bounded induction for the compositional truth predicate. *The Review of Symbolic Logic*, 2017. doi:10.1017/S1755020316000368.
- [47] Tin Lok Wong. Interpreting Weak König’s Lemma using the Arithmetized Completeness Theorem. *Proceedings of the American Mathematical Society*, 144(9):4021–4024, 2016.